DOCUMENT RESUME

ED 339 728 TM 017 631

TITLE	Computerized Placement Tests: Background Readings.
INSTITUTION	College Entrance Examination Board, Princeton, N.J.; Educational Testing Service, Princeton, N.J.
PUB DATE	Oct 91
NOTE	138p.; For related documents, see TM 017 416-417, and TM 017 630.
PUB TYPE	Collected Works - General (020) -- Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Adaptive Testing; College Entrance Examinations; College Students; Community Colleges; *Computer Assisted Testing; *Educational Assessment; Educational Technology; Guides; Higher Education; Item Response Theory; *Management Information Systems; Microcomputers; Pilot Projects; Predictive Measurement; Program Implementation; Research Reports; *Student Placement
IDENTIFIERS	*College Board Computerized Placement Tests

ABSTRACT

This document is a compilation of background readings for the user of Computerized Placement Tests (CPTs) developed by the College Board for student placement purposes. CPTs are computerized adaptive tests that test the individual abilities and backgrounds of examinees. CPTs are part of the ACCUPLACER student information management system. The first part of the compilation, "On Computer Adaptive Testing", contains the following papers: (1) "Using Microcomputers To Administer Tests" (W. C. Ward); (2) "On Item Response Theory and Computerized Adaptive Tests: The Coming Technological Revolution in Testing" (H. Wainer); (3) "Comparison of the Adaptive Test with Conventional Tests" (from the "Coordinator's Notebook for the Computerized Placement Tests"); (4) "An Overview of the College Board Computerized Placement Tests (CPTs)" (from the "Coordinator's Notebook..."); (5) "Summary of the Pilot Testing Results: 1984-1985" (W. C. Ward and others); and (6) "Test-Retest Results" (from the "Coordinator's Notebook..."). The "Implementing the CPTs" section contains the following papers: (7) "Computer Adaptive Assessment Testing at Central Piedmont Community College" (N. A. Webb and others); (8) "Uses of Computerized Adaptive Testing at Santa Fe Community College" (G. T. Delaino and others); (9) "The Computerized Adaptive Testing Program at Miami-Dade Community College-South Campus" (R. B. Schinoff and L. Steed); (10) "A Summary of Insights from Experience: Guidelines for Implementing Computerized Adaptive Testing for Assessment" (League for Innovation in the Community College); and (11) "Assessment's Next Wave: The Computerized Placement Tests" (P. Smittle). The third section, "Validating CPTs Results", contains the following papers: (12) "Statistical Analysis of the Class of 1994 at the USMMA" (C. Weber); (13) "Aims Community College Computerized Placement Tests Pilot Study, Summary Report" (D. Nold and M. Kuechenmeister); (14) "Validating CPT Reading Comprehension Test Standards: Employing Relevant College-Level Performance Criteria" (A. R. Napoli); and (15) "Correlations between CPTs Scores and Courses Grades" (M. Pomplun). A CPTs user list, as of October 1991, is included. (SLD)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
ARTHUR J. DOYLE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED339728

COMPUTERIZED PLACEMENT TESTS™

BACKGROUND READINGS

TM017631

COMPUTERIZED PLACEMENT TESTS™ is a component of the ACCUPLACER™ System.

BEST COPY AVAILABLE

COLLEGE BOARD

COMPUTERIZED PLACEMENT TESTS™

BACKGROUND READINGS

October 1991

Copyright© 1991 by the College Entrance Examination Board and Educational Testing Service.
All rights reserved.

College Board and the acorn logo are registered trademarks of the College Entrance Examination Board.

Computerized Placement Tests, CPTs, Seamless Testing, Serial Testing, and ACCUPLACER are trademarks of the College Entrance Examination Board.

Computerized Placement Tests™ is a component of the ACCUPLACER™ system.

Computerized Placement Management Software and the Placement Research Service are also components of the ACCUPLACER™ System.

Computerized Placement Tests™ (CPTs) were developed jointly by the College Board and Educational Testing Service.

Table of Contents

Introduction	v
On Computer Adaptive Testing	
Using Microcomputers to Administer Tests	1
On Item Response Theory and Computerized Adaptive Tests: The Coming Technological Revolution in Testing	9
Comparison of the Adaptive Test with Conventional Tests	19
An Overview of the College Board Computerized Placement Tests (CPTs)	21
Summary of Pilot Testing Results: 1984 -1985	23
Test-Retest Results	39
Implementing the CPTs	
Computer Adaptive Assessment Testing at Central Piedmont Community College	41
Uses of Computerized Adaptive Testing at Santa Fe Community College	51
The Computerized Adaptive Testing Program at Miami-Dade Community College	61
A Summary of Insights From Experience: Guidelines for Implementing Computerized Adaptive Testing for Assessment	77
Assessment's Next Wave: The Computerized Placement Tests	83
Validating CPTs Results	
Statistical Analysis of the Class of 1994 at the USMMA	89
Aims Community College Computerized Placement Tests Pilot Study, Summary Report	107
Validating CPT Reading Comprehension Test Standards: Employing Relevant College-Level Performance Criteria	123
Correlations between CPTs Scores and Course Grades	131
Computerized Placement Tests User List	135

INTRODUCTION

When I was invited to prepare the introduction to this resource-laden publication, the significance of my eight years of experience with the development of computerized adaptive testing and its subsequent application through the Computerized Placement Tests (CPTs) became apparent. Not only has an alternative model of assessment been developed, but its initial application for higher education has proven itself a landmark innovation.

In undertaking this assignment, my first task was devising a way to tell the full story. A second was how to recognize all those who contributed to making the CPTs more than a major advancement in measurement and technology by creating closer relationships among assessment, teaching, learning, and success for students and institutions alike.

Neither task can be fully completed here, but the articles in this publication do tell the story in a remarkable manner. Their authors reflect the breadth and depth of institutions and professions involved in this educational advancement and provide documentation for an important chapter in education that might otherwise become lost in an era of accelerating change.

For me, the greatest satisfaction has been witnessing the creativity of CPTs users. Their vision and openness to additional ways in which computerized adaptive tests can serve education has led to the evolution of a novel system supporting college entry-level assessment and placement. This experience has provided an incentive for pursuing companion models of computer-delivered assessment.

While I would like to acknowledge everyone who has made the delivery of adaptive testing and the success of the CPTs a reality, it is impossible here. However, the editor will permit me to do so symbolically, so I have chosen to mention Robert H. McCabe, president of Miami-Dade Community College and past chairman of the College Board. His sustained encouragement during the project's early years was invaluable.

Enjoy the readings ahead of you, and as you do think of sharing your experience with the CPTs and the ACCUPLACER system in the next edition of this publication.

Arthur J. Doyle
Executive Director
College-Level and State Services
The College Board

October 1991

Using Microcomputers to Administer Tests

by William C. Ward

The use of the microcomputer as test administrator has only begun to become feasible in educational settings. Until recently, machines have been too few and too expensive to justify their dedication to testing. Instead, initial applications have been made in other contexts such as personality testing in medical and psychological facilities. In these cases, the cost per examinee is not so important a consideration, and equipment needs are minimized because examinees typically present themselves one by one, rather than in the larger groups characteristic of educational testing.

The situation is changing in several respects. First, the availability of the needed equipment is increasing at a dramatic pace. A survey Creative Strategies International indicates that microcomputer sales are rising at an overall rate of 45% per year. In another survey Market Data Retrieval found that in a recent 12-month period the number of public schools using microcomputers for instruction increased by over 60%. Moreover, the cost of a given level of computing power is decreasing by approximately 30% per year, and the machines and their software are becoming increasingly easy for the nonexpert to use effectively. At least half a dozen colleges and universities have announced plans to acquire or to have their students purchase equipment that will result in the availability of computer resources to virtually any student at any time. Schools are investing heavily in the use of computers for instruction and for managing student records. In this context, paper and pencil testing eventually will appear as an antiquated impediment to processes that otherwise enjoy the speed and efficiency of electronic handling of information.

Equally important are the growing number of opportunities to use computers to augment what is possible in paper and pencil measurement. For example, computers can be used to improve the efficiency of measurement of familiar attributes, to obtain more detailed information about an examinee than is feasible in conventional measurement, and to assess new aspects of performance. We will consider several of these possibilities including the advances that are within the scope of current knowledge and those that hold promise for a more distant future. First, however, we will review some of the attractions of the microcomputer in administering tests.

It should be noted that little in this discussion relates specifically to the use of stand-alone microcomputers. Many postsecondary institutions have instead chosen computer systems in which a single central processor controls large networks of inexpensive terminals. Intermediate to these two options are systems in which a number of micros are linked with a single hard disk to increase their memory resources. Differences in speed, storage, and graphics capabilities associated with given hardware choices may, affect their suitability for particular testing applications, but in general the issues remain the same regardless of the equipment configurations used.

Use of the Microcomputer in Administering Conventional Tests

The potential benefits of administering conventional tests by computer range from opportunities to individualize assessment to increases in the efficiency and economy with which information can be manipulated. Some of these capabilities are primarily of interest in the context of classroom or local institutional testing, others apply to large-scale standardized testing.

Much of educational testing traditionally has been managed on a mass production basis. Logistical considerations have dictated that all examinees be tested at one time and within fixed time limits. Under such conditions, the speed with which an examinee completes test questions often becomes an important factor and confounding variable in the assessment of levels of skills.

The computer as test administrator offers an opportunity to manage assessment differently. Efficient use of the equipment calls for a more flexible scheduling; students can be allowed to take tests at virtually any time. In addition, there is no need to require that all examinees complete a test in the same number of minutes, unless speed is an integral part of the attribute assessed. The computer can record the time an examinee takes, either for an entire testing session or for the individual questions within it, so that speed of response can be measured. The test constructor then has the option to base scores on whatever combination of speed and power is appropriate to the situation. For special subgroups such as handicapped students, this option can result in test administrations that are more similar to those under which the majority of the population is tested than can be achieved with traditional testing methods.

Moreover, the computer makes it convenient to employ different assessment strategies, such as providing immediate feedback on the correctness of the response to each question, or allowing the examinee to continue responding until the correct answer is obtained. The former may be attractive for its potential instructional benefit; the latter takes advantage of the fact that examinees often have partial knowledge even when they are unable to narrow the alternatives to a single choice.

Efficiency in Manipulating Information

The benefits from computer administration of conventional tests may be most pronounced in the flexibility and efficiency it creates for managing test information. In some contexts, such as college placement testing, immediate generation and reporting of test results may be especially important. The computer can produce a score that will be reported to the examinee at the end of the testing session, and it can automatically generate a variety of individual and summary reports for faculty and administrators. Moreover, the report can easily incorporate decision rules that are unique to an individual academic setting. Thus, if eligibility for Mathematics 101 depends on a score of 25 or more on the Mathematics Placement Test, the report can include both the student's score and information about the course for which he or she should register.

Greater efficiency can be obtained if the testing procedure can be linked electronically with other aspects of an institution's registration, advisement, and placement system. Additional information about a student can be incorporated into the test interpretation and consequent decision making; tailored advice can be provided, without requiring that counselors be involved in the routine part of the advisement process; and decisions can be incorporated automatically into the student's record.

Economy

The economic value of individualized and automated manipulation of information will differ from situation to situation. When the examinee's time is scarce or valuable, as in professional licensing examinations, the flexibility of computerized testing may be an important benefit. When test results are needed quickly, as in placement decisions, the immediacy of score reporting may be especially valuable. In other settings the chief value may be an institution's ability to handle large volumes of student data with minimum time and effort required of professional staff.

For many schools and colleges, the costs of obtaining the equipment needed for computerized testing are currently greater than could be justified for testing alone. As microcomputers are acquired for other purposes, however, and particularly as they become central to instruction in many areas, the marginal costs of testing by computer will become more favorable. Furthermore, cost trends indicate that using computers in place of paper-based tests and records will require less personnel time. Industries such as banking already have obtained major economic benfits from investing in automation, and similar changes will eventually be seen in educational institutions.

Computerized Adaptive Testing

Computer administration of conventional tests may improve the efficiency with which information about an examinee may be handled. However, it does nothing to increase the quality of that information. Use of the computer to administer an adaptive test is one way technology can enhance the information obtained from a test.

An adaptive test is one in which examinees receive different test questions, according to what the individual's performance indicates as being most appropriate for his or her level of skill. There are various schemes for accomplishing this objective, but essentially the test proceeds as follows: One or more initial questions are selected for each examinee, perhaps by drawing randomly from the available pool of questions. After these questions have been answered, the computer generates a score representing the best current estimate of the examinee's ability, then selects from the pool the next question that is most appropriate in light of the ability estimate. The question is administered, ability is reestimated on the basis of previous performance, and the best question for the examinee is selected from those remaining in the pool. This procedure continues until the desired number of questions has been given.

Adaptive testing is made possible by an advance in measurement theory known as Item Response Theory (Lord, 1980) which provides the mathematical basis for selecting the appropriate question to give at each point and for producing scores that are comparable between individuals. Item Response Theory has been the subject of extensive theoretical and empirical research for 30 years. Publishers of standardized tests have successfully employed Item Response Theory in tasks such as equating one form of a test to another. Because of its substantial computational requirements, it has only recently become feasible for use in microcomputer testing, but it is now the subject of major development efforts by the military, test publishers, and computer hardware developers.

The major benefit adaptive testing offers is the opportunity to obtain efficient measurement over a broad range of ability. A conventional norm-referenced test is a compromise; ordinarily, it is composed mainly of moderately difficult questions for a specific population so as to provide accurate measurement for the majority of those with whom it will be used. Relatively few difficult questions, able to discriminate among individuals whose ability level is extremely high, or easy ones, to distinguish individuals with minimal skills, can be included. With adaptive testing, this compromise is not required. The pool of questions can include enough easy and difficult questions to provide good discrimination across the entire range of abilities for which the test is intended. At the same time, this range can be achieved with a test that is shorter than typical conventional tests. Individuals with low skills never encounter the difficult questions on which they would resort to blind guessing; others do not waste time with easy questions that they would almost certainly, answer correctly. Each individual confronts only the questions that are likely to provide substantial information at his or her level of skill.

Given a well-constructed pool of questions, the savings in the number of test questions required can be substantial. A conservative estimate would be that the length of a test battery can be cut by 50 to 60% and still maintain a measurement accuracy equivalent to that of the best of standardized conventional tests. Depending on the purposes of the assessment, this reduction can be used to save examinee time or to make available time for administering additional instruments, so as to increase the information on which a decision is based.

Attractive as its potential benefits are, adaptive testing is not for every assessment situation. Its computational requirements limit its implementation to the more powerful microcomputers commonly found in educational settings. Its mathematical assumptions also are demanding; in particular, models that have been thoroughly explored assume that the attribute measured by an adaptive test is unidimensional. This condition has posed no difficulty in practice for the assessment of broad aspects of developed abilities such as reading comprehension, but it requires careful examination in other contexts, including applications to achievement testing in specialized content domains. Moreover, extensive data are needed to prepare an appropriate pool of test questions; the more rigorous Item Response Theory models in

current use require that at least 1,000 examinees complete each question. Thus, professionally developed, large-scale standardized testing, such as admissions and placement testing and professional licensing and certification, provide the most promising opportunities for adaptive testing applications.

Diagnostic Testing

An adaptive test seeks to measure as efficiently as possible one or more broad dimensions of an individual's aptitudes or skills. A diagnostic test seeks to decompose these broad dimensions into specific components to identify in detail an individual's strengths and weaknesses. A major goal in diagnostic testing is to provide information instructors can use in determining what a student needs to learn and what skills he or she has on which this learning can be built. The test also should suggest to the teacher what instructional approaches are likely to prove effective with the student.

The ideal diagnostic test would incorporate a theory of knowledge and a theory of instruction. The theory of knowledge would identify the student's skills and the theory of instruction would suggest remedies for the student's weaknesses. Moreover, the test would be – in a different sense of the word from that we have previously employed – adaptive. That is, it would not subject students to detailed examination of aspects of skills in which they have acceptable overall competence or in which they are totally lacking, but would quickly focus on areas in which a student has important strengths and weaknesses – areas where an overall score is not an adequate representation of the individual's status.

Available instruments fall short of this ideal. Lacking strong theory, test constructors must rely on a combination of logical analysis of the domain to be assessed and instructors' perceptions of the kinds of scores that will be useful in their teaching. Exercises directed to the skills identified by such analyses can be provided, but it is up to the instructor to integrate these into an effective program.

For the present generation of diagnostic tests, information management may be the most important role for the computer. Whatever the basis for creating profiles of students' skills, diagnostic testing implies that many such skills will be identified, and effective use of this information requires specific instructional interventions. The classroom instructor may have difficulty tracking and responding to the needs of 25 students when little more than their overall level of ability in a field of study is known. When 10 or 20 or more scores per individual are available, the task approaches an impossibility. The computer can assume responsibility for the record keeping that will enable the instructor to monitor students' performance, the instructional sequences they have been given, and the outcomes of those interventions.

Advances in diagnostic testing are occurring on many college campuses, not in the name of testing but through the development of programs of computerized drill and practice to supplement classroom instruction. Recent conferences of college English instructors, for example, have been rife with reports of implementations of such programs. Testing receives little attention in these reports, not because it is absent, but because it is so thoroughly embedded in the instructional sequence. The objectives of a unit define the skills to be diagnosed. Each question a student responds to provides practice and a scorable bit of performance that can be used to determine when these objectives have been met.

Future developments in diagnostic testing are likely to include a continuation of this work, with instructional objectives paramount, with testing an integral but subsidiary adjunct, and with the role of the computer expanding to encompass all aspects of instruction that teachers can make routine under specified circumstances. At the same time, the computer will have a second role to play: assisting in research and development to arrive at a better theoretical understanding of the skills that constitute competence in a domain. As methods and concepts of cognitive science are applied to the study of these skills, they will lead to the development of models of competence that are far more detailed and more useful in diagnosis than are traditional analyses based on the surface attributes of test questions that students can and cannot answer correctly.

The work of Brown and Burton (1979) in diagnosing the bases for children's errors in elemen-

tary subtraction provides an example of this approach. Brown and Burton identified several hundred specific systematic misconceptions, or "bugs," underlying such errors; for example, a student might systematically subtract the smaller digit in each column from the larger, regardless of which is on top. Identification of errors with this level of precision offers the possibility of specific remediation, and the statement of the error leads directly to a prescription for the necessary instruction.

Even so sharply delimited a domain as subtraction gives rise to a highly complex analysis. Bugs are numerous, they interact rather than function independently of one another, and sometimes they change their surface manifestations from one testing session to another. It will not be easy to carry out comparable analyses in domains that are less well defined than elementary arithmetic. Nevertheless, these types of analyses may eventually lead to a new generation of assessment instruments that can be linked more directly to instructional sequences than are present tests. Because of the complexity of the models and their application to the analysis of a given student's performance, the computer will be an indispensable tool in these efforts.

Computerized Preparation for Taking Tests

Bookstore shelves have long contained an impressive number of publications designed to help students prepare for the standardized examinations for admissions into college and graduate and professional schools. Software catalogues offer a large assortment of programs with the same intent. One recent article reviewed 10 programs designed to prepare students for the Scholastic Aptitude Test alone (Silverman & Dunn, 1983). As do their printed counterparts, these programs range from those that provide mainly drill and practice to those that offer instruction and attempt to teach effective strategies for test taking. Some include supplementary workbooks, and many include a practice test that sometimes indicates the time constraints under which students will complete the real test. At least one advises the student on effective management of the time available within the testing setting.

There would be no point in repeating the debate about the value of special preparation for these tests. Everyone would agree that students should be familiar with the kinds of questions they will encounter and the time constraints under which they will work, and few would claim that a brief period of special instruction can serve as more than a supplement to the aptitudes and specific knowledge a student has already developed. Given the interest of students in preparing for these examinations and the effort schools make to help them do so, it is appropriate to ask instead what role the computer can play in the process.

One important benefit may be motivational. Although the evidence is largely anecdotal, there are numerous reports of students' positive response to computer instruction. Particularly striking are the reports of English instructors who find students eagerly attacking drill and practice exercises that they dislike in other settings. Test preparation is not inherently exciting, and the rewards it may offer are deferred; anything that enhances students' motivation may be valuable for that reason alone.

A second benefit, only partly realized in existing programs, is the way in which the computer can monitor the student's performance. The computer can administer practice tests under realistic time constraints, track the student's use of available time, and advise the student about managing time. It can branch back and forth between practice testing and instruction, depending on areas of strength and weakness identified in the testing, and it can systematically reintroduce questions or types of questions on which the student needs additional work. In short, it can provide a degree of individualized preparation beyond what is available in a classroom setting.

Finally, for schools that wish to offer test preparation, there may be economic benefits. Silverman and Dunn (1983) estimated that for their school, the costs of hardware and software acquisition would average $20 per student over a 5-year period whereas the cost of hiring a teacher for a classroom course would be eight times as great.

Computerized Simulations of Problem Situations

A final current application is in the presentation of the kinds of branching problem simulations that are popular in professional licensing and certification testing. The "patient management problem" found in medicine and other health-related fields provides a prototype. The examinee confronts a situation, makes an initial decision (for example, should the patient now be given laboratory, tests or be asked to provide a history?), and is branched to further decisions (Which of the following laboratory tests should be made? What do you do now that the patient's condition has worsened). The problem continues until the patient is cured, released from care, or, sometimes, killed.

These problems are attractive because they offer greater realism than do standard "one-shot" examination questions. and because they provide samples of performance that can be scored in multiple ways; for example, how efficiently did the examinee attack the problem, how many serious errors were made, what was the cost to the patient in dollars and in pain and suffering, and so on. In paper form, however, they are cumbersome and expensive to administer and to score. A test question can consist of a thick booklet, requiring the student to flip back and forth to appropriate pages, and latent image printing or a similar method must be used so that examinees cannot look ahead before their answers have been recorded. In scoring the performance, a method that allows the test evaluator to reconstruct the path taken by the examinee must be used because the value of a given action may depend on a particular point in the problem. A large volume of data from each problem must be handled to permit generation of the multiple scores that may be desired.

Computer administration of such problems has obvious advantages. For the test constructor, no special booklets or printing are needed; concerns about examinees looking ahead for clues are eliminated; and recording the sequence of decisions and generating multiple scores are automatic. For the examinee, the distractions of latent image markers and continual directions to turn from one page to another are eliminated. Thus, the cost of testing can be lessened and greater realism can be achieved.

For the Future

We have reviewed some of the benefits available from using computers to administer the kinds of tests that are most familiar in educational settings, and we have discussed four applications that represent current or near-term enhancements of assessment: computerized adaptive testing, diagnostic testing, test preparation, and administration of simulations of complex problem situations. Each of these applications offers one or more important improvements, including the individualization of assessment, increased accuracy, efficiency in handling test information, more detailed information to guide instruction, and decreased costs. These benefits alone are sufficient to guarantee that the role of the computer in test administration will become increasingly important, and that in many settings it will displace paper-based testing entirely.

The possibilities not yet realized, however, are more dramatic than the present applications. One basis for this claim comes from projections concerning technological developments. A continued multiplication of the computing power available at a given cost is only one aspect of the changes that can be expected. A similar revolution in optical disk technology may soon offer virtually unlimited cheap storage of information in a desktop machine, along with the ability to select from a large store of high-resolution visual displays. The range of test stimuli that can be presented will then be constrained only by the ingenuity of the test constructor. At the same time, advances in the ability to communicate among micros and between micros and larger machines will open the way to economical sharing of software and data base resources and to greater efficiency in handling test data through electronic means.

Advances in cognitive science and artificial intelligence will provide a major impetus in developing new methods of assessment. As suggested in the discussion of diagnostic testing, these fields will bring to testing the ability to understand an individual's knowledge and skills at a depth not now possible, and to design methods of instruction

that are sensitive to specific errors and deficiencies that at present are hardly detectable.

We cannot begin to describe the range of advances that may be expected in the next 5 to 15 years, and we have no reason to believe that anyone's guesses will prove accurate in hitting either the magnitude or the direction of the changes to occur. We will end this discussion by suggesting two illustrations of ways that tests of the future may differ from those of the present.

One potential development is in the nature of achievement tests. We know, from work in cognitive science, that experts in a field differ from novices in the information they possess, in the structure of their knowledge and in the way they bring that knowledge to bear in solving a problem. Chi, Feltovich, and Glaser (1981), for example, who studied the process of solution of problems in physics, showed that experts categorized and approached problems in terms of the principles required for solution ("It's a Second Law problem"), whereas novices addressed the same problems in terms of their surface features ("It's an inclined plane problem"). As an understanding of how to describe such differences is acquired, tests that assess more than isolated facts and principles can be developed, allowing an evaluation of the degree to which an individual thinks like an expert in a given domain.

A second illustration involves the assessment of an examinee's freely written responses. For some domains, where a given question has a limited number of acceptable answers and where key words or phrases identify those answers, the computer must only be provided with an exhaustive list of responses for which it is to scan. In such areas the computer can easily be used to allow testing to go beyond the multiple-choice question to questions that require production rather than recognition – a small matter, but one that will be attractive to many test takers and users.

More interesting and more difficult will be the use of the computer to assess the quality of a student's writing. From the perspective of instruction, there are good reasons for having students compose on the computer, particularly because of the ease with which writing can be edited and revised. Substantial progress in the analysis of writing samples has been made with systems such as Bell Laboratory's Writer's Workbench and IBM's Epistle; these systems are able to measure many lower-order attributes of a body of writing and to detect a variety of errors and weaknesses in usage. More powerful language parsers and dictionaries with richer networks of semantic information may be expected from the linguistic and artificial intelligence work now under way. We cannot yet see the day when evaluations of stylistic nuances of polished writing can be automated, but it is a reasonable hope that within the next decade we will have a microcomputer capable of grading the freshman essay.

References

Brown, J. S., & Burton, R. R. (1979). Diagnostic models for procedural bugs in basic mathematical skills. In Tyler, R. W., & White, S. H., (Chairs), *Testing, Teaching, and Learning.* Washington, D.C.: National Institute of Education.

Chi, M., Feltovich, P. J., & Glaser, R., (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science,* 5, 121-152.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum.

Silverman. S., & Dunn, S. (1983). Raising SAT scores: How one school did it. *Electronic Learning,* 2, 51-53.

William C. Ward

Director of Measurement Technology Research in the Research Division of the Educational Testing Service. He is currently directing a joint ETS/College Board project to develop a microcomputer-based adaptive testing system and to apply adaptive testing to basic skills placement testing at the college level.

Reprinted from the *Computerized Placement Tests: Coordinator's Notebook for the Computerized Placement Tests,* July 1986.

On Item Response Theory and Computerized Adaptive Tests: The Coming Technological Revolution in Testing

by Howard Wainer

Large-scale ability testing has been a conerstone of American education for most of the twentieth century. It has been relied on to facilitate university admissions decisions, to measure the output of schools, to evaluate the efficacy of educational interventions, and to serve the many other purposes that form the basis of the American meritocracy.

In the decade of the 1970s, the testing industry was subjected to substantial criticism. Some was deserved, but much was due to the critics' poor understanding of the complex technical issues involved. Clearly, some responsibility for this misunderstanding can be laid at the door of the testing industry, which did not take sufficient pains to explain these details until after the unfortunate criticisms had been made public.

We are now on the verge of a technological revolution in testing as a result of the availability of extensive inexpensive computing and some recent developments in statistical test theory. The purpose of this article is to try to explain accurately, but in nontechnical terms, the structure of this revolution and the role played in it by test theory.

One purpose of testing is to allow the user to compare different individuals on the traits or skills being tested. If we wish to make valid comparisons, there are only three possible ways to do so.

1. Use the same test for everyone. This is sensible in some circumstances; however, if we need to use the test more than once – for example, to make comparisons across time – questions of test security become important. (This is especially cogent with the new disclosure regulations.) Additionally, as we gain experience with various test questions, we often find ways of restating them to improve them – reducing ambiguity, removing misleading alternatives, and so forth. Last, changing social mores can alter radically our view of what are acceptable topics for testing. Thus, using a single form of a test is useful only for a limited set of circumstances. In most cases we need several forms, and they must be able to evolve.

2. Equate forms through common parts. We can still make comparisons among individuals who take different forms of a test if we construct the various forms so that they are, in some sense, equal. A common definition of successfully equated test forms is centered around the concept of "form indifference." That is, if the examinees knew how each of the forms was constructed and the statistical manipulations that would subsequently take place with the scores, they would be indifferent as to which form they were given. Methods of test equating have been developed to a very great degree of sophistication (see Angoff, 1957; Holland and Rubin, 1982). Most methods rely on the two forms that are being equated having a common portion. This allows the individuals to be compared directly on that subsection and the remaining parts to be equated from this knowledge. If a test form is compiled separately for each examinee in order to be, in some sense, ideal for that individual, there may not be any portions in common. Thus, the usefulness of these methods would be reduced.

3. Connect different forms with theory. If a person's height is measured in California using one ruler, and another's is measured in New York with a different ruler, on what basis can we state who is taller? What connects the two rulers? The answer is theory. Underlying our measurement of height is a theory of physical length that allows us

to manufacture rulers according to standards. So we never need to place any two rulers side by side to know (within reasonable bounds of error) that measurements made on these two different rulers are comparable. While this analogy is not completely accurate, in that it contains parts of test equating as well, it will serve to illustrate the points I want to make next.

It is well established that a good test should center the difficulty of its items around the ability of the person taking it. We learn little by asking questions that are much too easy; they only bore the examinee. Similarly, we do little more than frustrate the examinee if we ask questions that are much too difficult. Thus, tests given to fourth graders to measure reading ability are quite different than those given to twelfth graders to measure the same ability. The rulers are of different lengths, although the measurements can be compared.

When it is necessary to give a test to an examinee pool of broad ability range, we must have questions on the test across the entire range. A test constructed under these circumstances usually has more items in the middle of the ability range and fewer at the extremes, but this is a compromise at best, aimed at boring and frustrating as little as possible. Further, items are often arranged in ascending order of difficulty, so that the less able students have answered all the questions that they plausibly can before hitting the difficult ones that might frustrate and anger them. The most able students, sadly, must wade through the too-easy items.

The coming technological revolution in testing is the Computerized Adaptive Test (CAT), an impressive amalgam of high-powered inexpensive computing and theoretical developments. Statistical test theory, in fact, provides the means of building cohesive, meaningful tests from items contained in the computer. I shall now embark on an extended metaphor that I hope will explain and illustrate CAT and modem Item Response Theory (IRT) – and how they fit together to yield a powerful new measuring instrument that will permit testing with greater accuracy and efficiency and less boredom and/or frustration for the examinees. It will also reduce the possibilities for cheating, provide information about the appropriateness of the test for those being tested, and open new avenues for ability testing.

The Metaphor

"while you and I have lips and voices which are for kissing and to sing with who cares if some one-eyed son-of-a-bitch invents an instrument to measure spring with?"
e.e. cummings, 1926

In a nearby laboratory, there are two groups of researchers. The first group is in complete sympathy with the notions expressed by Cummings. The second also enjoys kissing and singing, but feels that it is sometimes important to measure certain aspects of spring. They are currently engaged in identical research programs, namely the study of hurdle-jumping ability in humans. The strategies these two groups use are quite different, and it is instructive to consider them.

The first group had panels of expert coaches studying the movements and builds of various subjects. They had them running and jumping, and they kept copious notes. Later the notes were compared and attempts were made to arrive at a consensus about each subject.

The second group had a long runway constructed with a sequence of hurdles spaced out along it; the hurdles started out very low and gradually got higher and higher. Each subject was instructed to run along the runway and jump over each hurdle as it came. Most would get over the lowest hurdles easily, and would knock over the highest ones. A clerk recorded the pattern of hurdles (standing or knocked over) after each runner completed the course. The runner's hurdling ability was somehow tied to the height of the hurdles that were successfully negotiated.

The directors of these two projects were questioned regarding their aims and the probable outcomes. It turned out that this was but one part of a much larger enterprise. Similar sites are to be set up all over the world, and careful records are to be kept of hurdling ability so as to monitor individual improvement and note any changes that

might occur in the general hurdling ability of people over time. The director of the first project said his group was training experts to make judgments about hurdling ability based upon in-depth study of each runner. Asked how the experts would measure change in hurdling ability over time both within the same individual and among all children of the same age, he replied that a good coach could make such judgments accurately.

The second study was quite different. This group had a mimeograph machine working at top speed duplicating an instruction sheet that specified the number of hurdles, the distance between them, the height of each hurdle, their order, the kind of track they must be on, allowable bounds of temperature, wind, etc., etc. The researchers also were preparing packages of hurdles to be mailed to the other testing sites.

These hurdling tests embody much of what we think is sensible about the measurement of human ability. They also present a context for the discussion of testing and test theories past, present, and future.

Often what we are interested in when we give a test is a measure of the ability of the examinee. In hurdling tests, the measure of ability relates to the heights of the hurdles cleared. This can be made operational in a variety of ways. If "1" represents a hurdle cleared and "O" one knocked over, someone performing as expected will have a response pattern like this one: 1111110000. If the hurdles are in order of increasing size, we could then characterize a person by the height between the highest hurdle cleared and the lowest one missed. For example:

Height(cm)
10 20 30 40 50 60 70 80 90 100 110

Performance
1 1 1 1 1 0 0 0 0 0 0

This pattern indicates the subject's ability is about 55 cm.

The second question we would ask is the accuracy of this estimate. In this situation we might say that we were accurate to within 5 cm (i.e., between 50 and 60 cm). Of course, this presupposes tremendous consistency on the part of the hurdler, but if the same result occurred each time the hurdler "took the test," we would arrive at the same conclusion, and our confidence in it would increase.

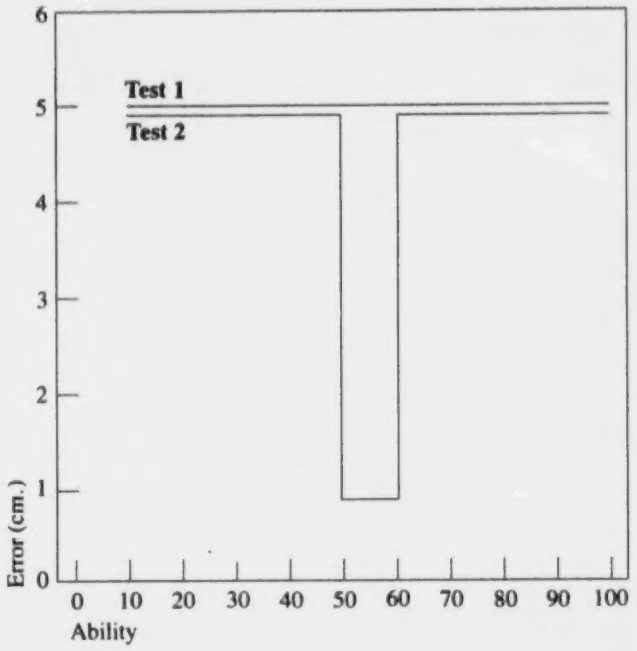
If we wanted more accuracy in the estimate, we would have to insert more hurdles between 50 and 60 cm for this person. This is a common problem in testing, for the increased accuracy obtained with a more finely graduated test necessitates an increase in labor for the examinee (to measure to the nearest centimeter, we would need 110 hurdles rather than the 11 used here). If all hurdlers have to attempt all hurdles, the increased labor is of little help for an individual whose ability is far from the heights being used; having successfully negotiated 10 hurdles between 10 and 20 cm does not tell much more about the 50 cm hurdler than noting that he or she cleared the 10, 20, and 30 cm hurdles.

Traditional practice is to increase the number of hurdles in the area where the greatest discrimination is required. Often we can get a feeling for the structure of the test by drawing a graph that indicates the error of measurement at each jumping level. The table (on the following page) shows a plot of the error for the test as it is now made up, and another for one in which there were 10 hurdles (items) inserted between hurdles 50 and 60 (yielding errors of .5 cm).

Suppose the same jumper runs down the runway again, this time clearing the 60cm hurdle but knocking over the 50cm. This tells us that his or her ability is still likely to be in the 50-60 range, but the error range is expanded a bit. This points out two components of error in ability estimation—the accuracy limitations of the test construction, and the variability of human performance. We can control the first but not the second.

Enough concepts have now been illustrated to permit comparing these aspects of testing as they operate in traditional test theory (e.g., Gulliksen, 1950), modern Item Response Theory(IRT), and in a test of the near future.

The error functions for two tests



Variability

Traditional test theory expresses the examinee's ability in terms of number right. The problem is that, although this approach works well when a person performs as he or she ought, it offers few obvious clues when something is amiss. For example, a hurdler who gets a response vector of 1111100000 would have a score of 5, as would someone with a vector of 1111000001. The latter response seems clearly erroneous, and one would suspect that there was either a clerical error in recording or the runner had sidestepped the last hurdle (corresponding to successful guessing on a test). It would also give a score of 5 to someone who scores 0000011111. With a response pattern like this we clearly ought to suspect something peculiar is going on with the person taking the test, and a sensible result ought to be to require a retesting. But with traditional test theory a score of 5 is a score of 5, and that's it..

A second shortcoming is that the ability scale derived is only ordinal. That is, if one person scores 4 and another 5, we perceive that the difference between them is the same as that between two persons whose scored abilities are 9 and 10. Clearly, changes in raw score do not have the same meaning all along the scale, even if the items are all evenly spaced. In addition, suppose the test had been expanded and there were 10 items between 50 and 60 cm; would an increase in score from 14 (hurdles 10 cm - 59 cm) to a score of 15 (jumping over all hurdles from 10 cm through 60 cm) mean the same as an increase from 15 to 16 (clearing 70 cm)? Of course not!

IRT, on the other hand, uses a function of the difficulty of the items answered correctly as an estimate of ability. Further, it yields a goodness-of-fit test that will clearly indicate when an unusual response pattern appears. Thus, the ability given to someone who jumps over all hurdles up to and including 60 cm will be essentially the same

regardless of how many hurdles were intervening between the 50 and 60 cm hurdles. The only difference will be the error estimate that is obtained.

It seems likely that ability estimates in the future will have essentially the same structure as those of current IRT, except that they will be directly referenced to material. In the case of the hurdling example, this means that a person's ability is directly related to the height of the hurdles jumped; thus, even if no one else ever took the test (how would one compute percentiles?), the results are valid and of interest – a 70cm hurdler is a 70cm hurdler no matter what else happens – and we can measure progress in a metric that makes good sense.

CAT would use the scoring methods of IRT, but would choose the heights of the hurdles to be jumped as the runner progressed down the path. Thus, an intermediate hurdle might be the first one. If successfully traversed, another, somewhat higher one would suddenly appear as the next to be attempted. If it was missed, a somewhat shorter one would be inserted. All of this would be done lickety-split so as not to interfere with the examinee's momentum. The need for a computer that processes millions of operations per second is obvious. The advantage of such a scheme over a long string of (mostly) uninformative hurdles is also clear. The computer can quickly zero in on the level of ability of the hurdler, then continue to give new hurdles until ability is determined to a specified accuracy.

To illustrate how this sort of scheme would work, picture a hurdler with a 50 cm hurdle. If it were cleared, the next one would be 75 cm. If it were missed, one of 62.5 cm would follow. If that one were cleared, one of 68.75 cm would be given; and if that were missed. one of 65.625 cm. And so on. The distance between what was passed and what was failed would be continually halved until the ability of the individual was estimated with acceptable accuracy.

Note that in this example the hurdler faced only five hurdles, and his ability was estimated within 3 cm. This would have been the case for anyone whose ability lay in a range 0 to 100 cm. To have accomplished this with a conventional test, we would have required hurdles every 6 cm, and the hurdler would have faced 11 hurdles before missing. Thus, the length of the test has been effectively halved, and for greater accuracy the savings would have been greater.

A further advantage is that the examinee who behaves in a regular way gets a very short test and can leave. Someone who behaves in an irregular way stays longer. Thus, the length of the test required to obtain a fixed amount of accuracy varies with the examinee. If a test is inappropriate for a particular examinee, this procedure will not converse within a reasonable amount of time, and the tester will be cued to the problem.

Accuracy of Measurement

Traditional test theory measures the accuracy of an estimate of ability by looking at how the test orders a group of people who have taken the test twice (or manufacturing two versions of the test artificially, using, say, odd and even items). If a test orders the people essentially the same way on two testings, it is said to be "reliable." The extent to which it does this is its "reliability."

Let us look at the components of reliability. First, it depends upon the discriminating ability of the items. (If the test had only two heights of hurdles, 20 cm or 20 feet, we would find that the test was not particularly reliable, because virtually everyone would perform the same way on it.) Second, reliability depends upon the inherent variability of the individuals being tested. (If on one administration a person cleared every hurdle and on the next administration cleared none we would be hard pressed to assign an ability estimate.)

Third, it depends upon the variability of ability in the sample being tested. (If everyone had the same ability, the ordering from one administration to the next would vary enormously, thus indicating an unreliable test, when in fact, the test could be quite good.) This problem is one of the gravest with traditional test theory – its reliance on reliability. It can make the character of the test appear to change with changes in the population being tested. Should the accuracy of a scale

depend upon who else was being weighed the day you stepped on it?

The effect of the norming group on test reliability and validity is not merely a statistical curiosity. It occurs often, and can stir up trouble when it is not understood. For example, it is common to find among first-year law students that the correlation between the students' Law School Admission Test (LSAT) scores and their undergraduate grade-point averages (UGPA) is zero, or even negative. This fact is sometimes used as evidence against the validity of the LSAT. The actual reason is that the LSAT was used – properly – as an admission device along with UGPA. This makes the admitted group much more homogeneous and so reduces the correlation between the measures.

One can understand why the correlation goes negative by thinking about the bivariate distribution. Some students will do poorly on both measures and, therefore, not be admitted. Others will do well on both – they go to Harvard. Thus, the students who attend most law schools and make up the majority of the sample have done relatively better on one than on the other, and so were admitted. Surely, a measure of a test's efficacy should not depend on who is being measured.

IRT deals with estimating the accuracy of a test through the standard error of estimate. This is essentially a function of the item structure. Thus, if we have items spread over 10 cm at one part of the test, a person whose ability falls in that part will be accurate to within 5 cm. If in another part of the test we have hurdles every centimeter, then the error at that part of the test is of the order of .5 cm. This is so regardless of the variability of ability among those taking the test. In fact, there need only be one person taking the test.

One problem is that it does not take into account the variability of the person taking the test (actually it does, but not as an individual, only on average).

Thus, the error estimate is sometimes a bit on the optimistic side.

In the future, test theory may be expected to expand or contract the error estimate currently in use on IRTs by the variability of the person taking the test, therefore yielding a more honest estimate of variability, but nonetheless one that is still independent of the group taking the test. Thus, we ought to be able to give a better estimate for a person who responds 1111100000 than one who responds 1111010000, or 1101001010.

Item Difficulty

In traditional test theory, difficulty is not well defined. In fact, in Gulliksen's classic text (1950), it is not listed in the index. A careful search turns up (p. 367 ff) a variety of definitions that relate the difficulty of an item to the proportion of individuals that get an item correct within a particular sample. This is the grist that will eventually be made into a measure of difficulty, but at the time it was merely one way of doing it. One problem with this formulation is that difficulty changes from one norming group to another. A more serious problem is that the concept of difficulty is not functionally tied to the concept of ability.

By contrast, the most fundamental concept of IRT is the functional relationship between the ability of a person and the difficulty of an item. Referring back to the hurdling test, this means that we can describe the likelihood of a person successfully negotiating a hurdle of a particular height as a function of the difference between his or her jumping ability and the height of the hurdle.

Difficulty is defined as a function of the proportion of individuals who answer an item correctly. A hurdle that is almost never scaled is called "difficult." One that is almost never toppled is considered "easy." These same designations are possible with traditional test theory, but IRT defines "difficulty" quantitatively and unifies it with ability within the context of a theoretical structure.

There is an apparent circularity here that needs to be explained. Item difficulty is defined by the proportion of people who answer that item correctly. Person ability is defined by the proportion of items that a person answers correctly. Yet I stated that IRT was relatively unrelated to such

things as norming groups, and that the accuracy of estimate didn't depend upon who else took the test. How does this follow?

The critical concept here is one of the difference between ability and difficulty being the variable of interest. Suppose we gather a group of individuals to take the hurdling test. We have no idea of the difficulty of the items or the ability of the people. Quickly, we find that some hurdles are easier to jump than others, and some people are more skilled jumpers.

Through the intervention of the IRT model, we obtain ability estimates for the people and difficulties for the items. Of course, they are not correct with respect to origin, but they are correct relative to one another. If we now give the same items to another group of subjects, we can calculate their ability on the same scale as the first group. Or we can have the same group jump different hurdles and calibrate those hurdles on the same scale as the original ones.

Note that we can choose a subset of the original hurdles to measure some new group of people and do it on the same scale. Then, if it is suspected that a particular group of new examinees may have greater ability (for hurdling, they may be taller or older), they might be given fewer short hurdles and more high ones (thus increasing the accuracy of the test in the area of their anticipated ability). This aspect of IRT is called "item-independent person measurement"– being able to measure individuals on the same scale in a way that is independent of the precise set of items chosen. Thus, items can be chosen in such a way as to reduce error of measurement.

In addition, because the items are calibrated by the difference between their difficulty and the ability of the norming group, item difficulty can be estimated from any group, if items are matched with people so that they are reasonably suitable. IRT can give us protection from ignorance, but not from stupidity. The characteristic of IRT that allows us to calibrate items on any group of individuals is called "sample-free item calibration."

The shortcoming of the estimation of difficulty within the context of IRT is that it is operationally related to the people taking the test. If, somehow, it could be separated from them and assessed independently, it would permit much more powerful conclusions.

Consider the hurdling test again. Suppose we kept track of how often each hurdle was successfully jumped, and then we began to make careful physical measurements of the hurdles themselves – their height and color, the distance between them, where they occur in the test, and so forth. Suppose we then tried to correlate these physical properties with the observed difficulties. It might be that we would find we could predict the difficulty of a hurdle. If so, we could then produce a new hurdle and, before anyone had actually tried it, be able to predict those individuals who would or would not be successful in jumping it. Obviously, in this example height is the crucial variable, and we have faith that, if someone's hurdling ability is estimated to be 60 cm plus or minus 5 and we present him or her with a hurdle of 47.3 cm, we can be pretty well assured that his or her chances of successfully negotiating it are greater than those of someone whose ability was measured to be 50 cm.

How much greater, and precisely what is each person's probability of clearing this untried hurdle, requires the actual use of the IRT model. The use of expert judges might also enable qualitative judgments to be made; it is the precise statements of the likelihood of success, and the error bounds around these statements, that are the strength of the mathematical model.

The preceding metaphor has illustrated the logic and the power of Item Response Theory, although it has left unexplained the Byzantine mathematical windings behind its actual operation. Since the development of IRT has been an active research topic for more than 50 years, perhaps a short chronicle of its history is in order. Its earliest incarnation is in the work of Thurstone in the late 1920s and in that of Guttman, Lawley, and Lord in the 1940s. Much of this work was of more theoretical than practical interest at the time, because the numerical implementation of the new theory was beyond the contemporary computational capability for any problem of realistic proportions.

Although the theoretical work on IRT continued in fits and starts, the work in test theory that had the greatest influence at the time was Guilliksen's Theory of Mental Tests (1950). It provided the backbone of classical true score theory, which contained some of the notions of Item Response Theory but none of its mathematical formalizations. A decade later Rasch published his monograph on the simplest IRT. This model could be approximated, fitted, and its parameters estimated and used with the primitive computing facilities then available. Thus, by the early 1970s, IRT had gained a loyal cadre of proponents (see especially the work of Anderson, Fisher, and Wright). Their work on the development of both theoretical and practical aspects of the Rasch model has provided a clear picture of the potential of IRT for improving the quality of modern psychological measurement.

As computing facilities improved, the more complex models could be used. This accelerated the work being done on them. One by one, the major stumbling blocks in the way of their implementation were surmounted. Exemplary in this regard is the scholarly work of Bock, Lord, and Samejima. We have now reached the point where several decades of scholarship on what initially appeared to be "pure/ basic" research are providing the theoretical foundation for the next major advance in mental testing – CAT.

Setting aside analogies and similes, let me briefly discuss the advantages of CAT over conventional testing; how it will change the way we give (and take) tests, and how IRT meshes with it to allow these things to happen.

Advantages of CAT

Adaptive testing's logic – tailoring a test to suit the individual being tested – harks back to the initial Binet intelligence tests, which were individually administered. Modern mass testing made such individualized attention impractical. Interactive computers can now replace the human test administrator; indeed, the computer can make decisions faster and more accurately than any human. This, coupled with the development of IRT, makes individualized testing attractive again.

There are many gains to be had through the use of such a measure, and we are on the edge of making these gains. I will enumerate some; for further details, the interested reader is referred to Bert Green's (1983) essay, from which most of the following section was abstracted.

- Test security is improved, to the extent that a test is safer in a computer file than in a desk drawer.
- An individual can work at his or her own pace, and the speed of response can be used as additional information in assessing ability.
- Each examinee stays busy productively – everyone is challenged but not discouraged.
- The physical problems of answer sheets are solved. No longer will a person's score be compromised because the truck carrying the answer sheet was overturned in a flash flood – or other such calamity. There will be no more ambiguity about erasures, no problems with response altenatives being marked unwittingly.
- The test can be scored immediately, providing immediate feedback for the student. This also has implications for using tests diagnostically.
- Pre-testing items can be easily accomplished by having the computer slip new items unobtrusively into the sequence.
- Faulty items can be immediately expunged, and an allowance for examinee questioning can be made.
- A greater variety of questions can be included in the test builder's tool kit. The multiple-choice format need not be adhered to completely – numerical answers to arithmetic problems can just be typed in. Memory can be tested by use of succes-

sive frames. With voice synthesizers we can include a spelling test, as well as tests of the understanding of spoken language. Video clips showing situations can replace longwinded explanations on police or firefighter exams.

Changing Practice

Currently, large-scale testing is done in one of two different modes. Educational testing, as in college admissions, is usually done on relatively few occasions with many people taking the test at once. Military and governmental testing often takes place continuously, with applicants showing up in small numbers in an ad hoc (but statistically predictable) manner. For paper-and-pencil tests, it is best to have a few large administrations. For a CAT, where each examinee will need his or her own computer terminal, such a method is wasteful. It isn't sensible to have many terminals, which sit idle most of the time, to be used only for these few administrations. It is far better to adapt to the continuous testing mode, in which examinees come in to a testing office, sit down in front of a terminal, and complete the test.

Such procedures are well suited to the current practice in industry and government. Educational testing will have to adapt.

Some fears have been expressed, usually by older people, that examinees might be intimidated by facing a computer. I suspect that this is unfounded. Anyone who has seen the close symbiotic relationship between machines and users in the video arcades would be sanguine about this as well.

The Role of IRT

Comparisons among individuals who have taken different forms of a test must rely heavily on a connective theoretical structure. Aside from a few orienting items at the beginning of a test, each CAT form is individually tailored to the person taking the test. Yet, at the end of the test we will have to assign an ability score (or scores, if the test is multi-dimensional) to each person. These scores are of little value if comparative decisions cannot be validly made. Item Response Theory provides this glue. It measures the hurdles: assessing the empirical difficulty of each item. It adds up the performance: comparing the items correctly responded to with their difficulty and estimating ability. It provides a stopping point: estimating the accuracy of the ability determination and so deciding whether the test is over.

IRT also decides on the structure of the test. Performance on initial items provides evidence of examinee ability that indicates what the subsequent items should be. For maximal efficiency, items that provide the most information are generally used. This happily coincides with those items that are challenging without being frustrating. Last, IRT provides a yardstick to measure the quality of the items, and hence the test.

Conclusions

We are now on the verge of a technological revolution in testing. Paradoxically, the new testing is, in a sense, a return to old-fashioned individualized examinations. The merits of a one-on-one examination individually tailored to each student's knowledge and ability will be once again achieved. Now, however, the arbitrariness and lack of objectivity of such exams will have been removed. Further, we will expand greatly the range of questions that the "examiner" can ask through the inclusion of ancillary audio and video materials.

This revolution, which combines the objectivity and efficiency that currently exist in large-scale paper-and-pencil tests with attention to individual differences, is powered by the availability of cheap computing and the development of a powerful theoretical structure of testing. The happy confluence of these two events makes this revolution possible—the practice of testing, and our lives, to the extent that we depend upon testing, will be much the better for it.

References

Angoff, W. H., "The 'equating' of Nonparallel Tests," *Journal of Experimental Education*, *25*, 241-247, 1957.

Bock, R.D. and W. Fitzgerald, "The B-F Spacial Visualization Test" (unpublished), 1972.

Green, Bert F., "Notes on the Efficacy of Tailored Tests," *Principles of Modern Psychological Measurement* (H. Wainer & S. Messick Eds.), Lawrence Erlbaum Associates (Hillsdale, N.J.), 1983.

Gulliksen, H., *Theory of Mental Tests,* John Wiley & Sons (New York), 1950.

Holland, P.W. and D. B. Rubin, *Test Equating,* Academic Press (New York), 1982.

Lord, F.M., and M.L. Novick, *Statistical Theories of Mental Test Scores,* Addison-Wesley (New York), 1968.

Lord F.M., *Applications of Item Response Theory to Practical Testing Problems*, Lawrence Erlbaum Associates (New York), 1980.

Rasch, G., *Probabilistic Models for Some Intelligence and Attainment Tests,* Nielson and Lydiche (Copenhagen), 1960. (Republished by The University of Chicago Press, 1980.)

Thurston, L.L. and E.J. Chave, *The Measurement of Attitude,* The University of Chicago Press(Chicago), 1929.

Wainer, H., "A Test of Graphicacy in Children," *Applied Psychological Measurement, 4,* 331-340, 1980.

Wright, B.D. and M. Stone. *Best Test Design,* MESA Press (Chicago), 1979.

Howard Wainer
Senior Research Scientist, Educational Testing Service, Princeton, New Jersey

Reprinted with permission from *The Journal of College Admissions.*

Comparison of the Adaptive Test with Conventional Tests

A number of comparisons can be made between the adaptive test of basic skills and conventional tests intended to measure the same clusters of skills. Work completed thus far permits three such comparisons, relating to (1) the accuracy with which skills are measured; (2) the equivalence of levels of scores derived; and (3) the correlation of scores across the two modes of testing.

Information on the accuracy of measurement was obtained through simulation studies. The four modules of the adaptive test yield scores with reliabilities around .90, as high or slightly higher than reliabilities of conventional tests requiring more than twice as many questions. Test-retest correlations, a second method of estimating reliability, also yield estimates around .90 for the adaptive test modules.

Data gathered in pilot tests with three institutions in three states provide the basis for two further comparisons. Many students who were given the adaptive test also completed one short paper and pencil test, which was constructed from the same pool of questions as the adaptive test but share no questions with it. Two-hundred thirty-four took a 17-question reading comprehension test, while 88 completed a 12-question test of elementary algebra. Performance on the paper tests was scored by Item Response Theory methods and converted to the Total Right score scale used with the adaptive test; this scale ranges from chance-level performance to a maximum of 120.

Score levels for the adaptive test and corresponding paper tests were very similar, differing by 1.6 Total Right score points for reading comprehension and 3.8 points for elementary algebra. These differences amount, respectively, to .07 and .13 standard deviations relative to performance on the adaptive test modules, and do not approach statistical significance ($t < 1$ in both instances). The correlations between corresponding scores were .79 for reading comprehension and .88 for algebra. The true score correlations – that is, the correlations that would be expected if both tests could be made perfectly reliable – were .93 and 1.0, respectively. Thus, correcting for the reliabilities of the tests, the relationships between the paper and computerized tests are nearly perfect.

In brief, the adaptive test yields accurate and efficient measurement; the adaptive testing process and the administration of the test by computer appear to have little to no effect on the nature of the skills that determine test performance or the level of skills an individual is estimated to possess.

Reprinted with permission from *Coordinator's Notebook for the Computerized Placement Tests*, July 1986.

Test Results

	Reading Comprehension	Elementary Algebra
Adaptive Test		
Mean	77.3	67.1
S.D.	21.8	29.9
Paper Test		
Mean	75.7	70.9
S.D	.23.2	29.3
t (Correlated Means)	0.11	0.27
Correlation Uncorrected For Perfectly Reliable Tests	.79	.88

An Overview of the College Board Computerized Placement Tests (CPTs)

Educational Testing Service and the College Board have worked together to develop the application of computerized adaptive testing to practical measurement situations. Adaptive testing uses modern statistical methods to tailor the difficulty of the test to the skills of individual examinees. In doing so, adaptive testing provides very efficient measurement across a broader range of levels of skills than can be achieved with conventional paper-and-pencil tests. Using the computer makes possible a rapid determination of which questions should be given to an individual; it also permits immediate reporting of the test results.

The first application of this process is the College Board Computerized Placement Tests (CPTs), a test of skills in college reading, English, and mathematics. The test will enable postsecondary institutions to ascertain whether students are prepared for college-level work or need remedial studies.

The test consists of five independent modules: Reading Comprehension, Sentence Skills, Arithmetic, Elementary Algebra, and College-Level Mathematics. Consistent with current emphases in instruction, content specifications have been designed to emphasize question types that require reasoning and problem solving.

ETS/College Board experience in developing and administering the New Jersey College Basic Skills Placement Test and the extensive data on placement practices collected by the New Jersey Basic Skills Council provide the basis for determining what levels of skills should be included. The test is intended to yield accurate measurement from the lowest to the highest level of skills that any of the colleges in the New Jersey program now use for setting cut scores for basic skills placement decisions.

About 200 high schools and 60 colleges participated in pretesting new questions written for these tests. Analyses of their data provided the basis for selection of the pools of questions to be included in the adaptive test.

Simulation studies show that the test provides very accurate measurement with far fewer questions than are required in typical conventional test batteries. The various modules require from 12 to 20 questions to produce scores with a reliability of .90 or better; well-designed conventional tests employ more than twice this number of questions to approach comparable accuracy. Subsequent analysis has supported these high correlations.

Thus far [as of July 1986], the following institutions have participated in field testing of the College Board Computerized Placement Tests (CPTs).

- Anchorage Community College, Anchorage, AK
- Bergen Community College, Paramus, NJ
- Brazosport College, Lake Jackson, TX
- California State University, Long Beach, CA
- Central Piedmont Community College, Charlotte, NC
- Community College of Allegheny County, Pittsburgh, PA
- Delaware Technical & Community College, Dover, DE
- Evergreen Valley College, San Jose, CA
- George Mason University, Fairfax, VA
- Glassboro State College, Glassboro, NJ
- Herkimer County Community College, Herkimer, NY

- Illinois Central College, East Peoria, IL
- J. Sargeant Reynolds Community College, Richmond, VA
- Kirkwood Community College, Cedar Rapids, IA
- Lethbridge Community College, Lethbridge, Alberta, Canada
- Mansfield University, Mansfield, PA
- Miami-Dade County Community College, Miami, FL
- New Mexico State University, Grants Branch Community College, Grants, NM
- New York University, New York, NY
- Oakton Community College, Des Plaines, IL
- Scottsdale Community College, Scottsdale, AZ
- St. Louis Community College, St. Louis, MO
- University of New Mexico, Albuquerque, NM
- University of Virginia, Charlottesville, VA
- Yuba College, Marysville, CA

Repinted from the *Coordinator's Notebook for the Computerized Placement Tests*, July 1986.

Summary of Pilot Testing Results: 1984 -1985

by William C. Ward, Roberta G. Kline, and Jan Flaugher

Abstract

Seventeen two- and four-year institutions participated in the pilot testing of the College Board Computerized Placement Test during 1984 and 1985. Data were available for over 2,500 students who took one or more modules of the adaptive test; participating institutions planned to test at least 2,000 more by the end of 1985. In most institutions examinees were volunteers currently enrolled in English, reading, or mathematics courses.

Students taking part in the 1984 pilot completed a questionnaire describing their reactions, and many took a short paper and pencil test for comparison with the computerized test. In both years of pilot activity, students' grades in relevant courses and their scores on other tests used for placement were obtained when available.

Reactions of students and of faculty and staff were very favorable. Both groups found the computer easy to use; both liked especially the adaptation of the test to a student's level of skills and the ability to obtain immediate score reports.

Analyses of the relations between test scores and end-of-term grades were done separately by level of course (developmental or introductory college-level). Grades in single courses constitute a very fallible criterion, confounded by a number of factors. Nonetheless, moderate correlations with grades, with median values generally in the range from .30 to .40, were found for all categories of courses examined except for developmental mathematics courses; in the latter, the median correlations were over .50. These coefficients are similar in magnitude to those found in studies with professionally constructed paper and pencil test batteries.

A few samples provided data appropriate for a comparison of relations to grades for sex and ethnic subgroups. There was no evidence of differential prediction for these groups, but, with small sample sizes and volunteer examinees, these results are far from definitive.

No specific contrasts can be made of relations to grades of the computerized test and other placement tests given to pilot examinees; in most instances the other tests available were institution-specific, and the sample sizes were too small to permit confident generalizations about differences in the relations. Over all samples and courses for which data are available, the computerized test scores showed higher correlations with grades than did scores from paper and pencil tests in 59% of the possible contrasts.

Summary of Pilot Testing Results

Seventeen colleges and universities participated in the pilot testing of the College Board Computerized Placement Test during 1984 and 1985. This testing was meant to serve several functions:

(1) To provide an opportunity for participating institutions to become familiar with the adaptive testing process and to explore its applicability to their placement testing needs;

(2) To determine the reactions of students, faculty, and staff to the computerized placement test, and to identify any difficulties they might encounter in taking or administering the test; and

(3) To gather data relevant to the validity of the test as a measure of the basic skills needed for college work and of its comparability to paper and pencil tests intended to measure similar skills.

The purpose of this report is to summarize the results of the pilot testing, with emphasis on analyses related to the validity of the test.

Sample

Institutions participating in the pilot testing are listed in Table 1. Twelve of these are two-year public colleges, seven with medium-size enrollments, four with large enrollments, and one categorized as having a small enrollment. Of the five four-year institutions in the sample, three are large public universities, one is a large private university, and one is a small public university. (Information about length of program, size of enrollment and governance was taken from the College Board's *The College Handbook 1984-85.)*

In pilot activities during the spring of 1984, 426 students enrolled in three participating institutions were tested. College Board and ETS staff spent one to two weeks on campus working with faculty and staff in the data collection. In addition to the computerized test, students completed a paper test designed to parallel one module of the adaptive test, filled out a questionnaire dealing with their reactions to the test, and in some instances took part in a small group discussion of their experiences. Students' academic records were obtained for analyses of the relations of the adaptive test scores to other tests given for placement and to grades in the English, reading, and mathematics courses in which the students were enrolled.

Subsequently, fourteen new institutions joined in a continuation of pilot testing. Most began testing in the winter and spring of 1985. In this phase, the institutions organized and administered their pilot activities without on-site assistance from College Board and ETS staff.

They made their own choices as to which students would be tested, which of the four adaptive test modules would be used, and what options, such as the inclusion of institution-specific background questions, would be implemented. The only requirement for participation was an agreement to provide to project staff students' grades in relevant courses and their scores on available paper and pencil placement tests, and for pilot coordinators and their colleagues to complete questionnaires describing their reactions to the testing process.

More than 2,100 students were tested in the 1985 pilot. In most institutions test-takers were volunteers, although a few schools required that intact classes or designated students take the test. Most students were currently enrolled in relevant courses, although one pilot site used the computerized test as part of its regular placement process and two others used it for pre- and post-testing of students in special intervention programs for those academically at risk. Sample sizes ranged from fewer than 20 students to nearly 1,000. Excluding the samples at the extremes, the average number of students tested was about 80.

A number of participating institutions scheduled continuing pilot activity through the fall of 1985 and were expected to have tested at least 2,000 additional students by the end of 1985.

Altogether, test results for 2,539 students were available at the time this report was prepared. The numbers of students completing any one test module ranged from over 1,500 for Elementary Algebra to over 2,000 for Reading Comprehension. Sixty percent of those tested were females, and 58% identified themselves as White.

Data Collection

In the initial data collection, a standard test configuration was used in which students completed all four modules of the adaptive test and answered a constant set of background questions. In the subsequent testing, thirteen institutions administered both verbal and mathematics tests, although not all students were given all tests; two administered only verbal tests; and one, only mathematics. Three requested a branching rule, in which students receiving scores below a designated value in Arithmetic automatically were branched around the Elementary Algebra test.

The majority of the pilot sites tested on IBM equipment, primarily PCs but including XTs and Jrs; Compaqs and Zeniths were also used. The number of students tested at one time ranged from one to nineteen depending on the number of computers available. Institutions that had access to

computer laboratories or testing centers were able to test larger numbers of students simultaneously.

Student Reactions

Student reactions to the computerized test, based on responses to the questionnaires administered in the 1984 pilot, are summarized in Table 2. In general, students found the computer easy to use and the test to be of appropriate difficulty. They showed little sign of computer anxiety and overall preference, by a substantial margin, for testing by computer rather than paper and pencil.

The questionnaires included spaces in which students wrote free-response answers to questions concerning their reactions. Answers to these questions and comments given in small group discussions indicated very positive response to several aspects of the testing situation: the ease of computer use, the absence of time limits, and the fact that they did not have to write answers or grid ovals on answer sheets. The only frequent complaint concerned visual difficulties with the computer display: monitors were sometimes old and of low quality, and room lighting was not always designed to avoid glare. Some students also objected to being unable to return to questions previously unanswered.

Analyses were conducted to examine the relations between students characteristics and their reactions to the test. In general, there were not strong differences in reactions associated with background characteristics. For example, the few students who reported that English was their second language were less likely than native speakers to report that it was very easy to use the computer; but a majority of each group endorsed this statement.

A final set of analyses examined student reactions in relation to their test scores. Students receiving high scores on a test more often saw the test as about right in difficulty for them. High scorers tended to report being very much challenged to do their best and to have felt some time pressure; however, they were also more likely to report that the computer was very easy to use and the directions were very clear, and that nervousness did not prevent them from doing their best. They did not differ from low scorers in their prior experience in using a computer, but were more likely to report having programmed one.

Reactions of Faculty and Staff

Questionnaires were sent to each of the institutions participating in the 1985 pilot testing. Eight testing coordinators completed an extensive questionnaire which sought their reactions and descriptions of the conditions under which testing took place; 30 additional faculty and staff members who had an opportunity to try out the system completed a briefer form dealing primarily with their own experience and reactions to the test.

Both groups responded favorably to the test. They reported that they and their students found the directions clear and the computer easy to use; a strong majority believed that students would prefer a computerized test to a paper and pencil one. Among the aspects of the test they liked best were its adaptation to a student's level of skills and the immediate scoring and reporting of results. Among those liked least were the inability to return to change the answer to a previously administered question, an option not appropriate in adaptive testing, and the lack of diagnostic information. The latter need will be addressed by the diagnostic test of basic skills currently under development by the College Board and ETS.

Descriptive Results of Pilot Testing

The mean scores for all students tested in the 1984 and 1985 pilots are presented in Table 3. For Reading Comprehension, the mean score corresponds to the 46th percentile in a normative sample, described in the Coordinator's Notebook; for Sentence Skills, Arithmetic, and Elementary Algebra, respectively, the means correspond to the 33rd, 42nd, and 42nd percentiles.

Also shown in Table 3 are the means for the most able and least able samples of students tested. On Reading Comprehension, for example, the most able sample tested at any participating institution had a mean score corresponding to the 68th percentile in a normative sample, while the least

skilled had a mean corresponding to the 19th percentile. In no case was any of the test modules too difficult or too easy for the typical student tested by an institution.

Table 4 displays median correlations among the test modules across all samples tested. As would be expected, the two mathematics tests and the two verbal tests have the highest intercorrelations. All the tests, however, are moderately correlated with one another.

Relations of the Computerized Test to Course Grades

Analyses of relations between test scores and end-of-term grades were done separately by level of course (developmental or introductory college-level); combining students across levels would have introduced a source of considerable extraneous variation into the relations. Correlations are reported whenever a group of 20 or more students was available. For this analysis, grades were coded using the nine-point scale shown in Table 5.

It should be noted that there is sometimes a severe limit on the magnitude of correlations that can be expected from these analyses. Grades in single courses are not a very reliable criterion under the best of circumstances. Moreover, in many instances it was necessary to combine results for students enrolled in a number of different courses so that a sufficient number of cases would be available for analysis; there may, for example, have been as many as half a dozen courses all classified as developmental English. To the degree that different courses entail different degrees of rigor, either because of the course objectives or because of differences in instructors' grading practices, the correlations will be adversely affected.

Further, in several instances the distribution of grades was so restricted in range that appreciable correlations would be very unlikely. For example, in one institution the grades given in developmental reading ranged only from Satisfactory to Unsatisfactory, and more than 70% of all those graded received a grade of Satisfactory. Finally, to the degree that students assigned to a given level of course work have been appropriately placed for their level of skills, the computer test scores will also be restricted in range.

Despite these factors, average results across all the samples available for analysis show very respectable relations between grades and test scores. The results are displayed in Tables 6-10 and summarized below.

Developmental English

Eight institutions tested a sufficient number of students enrolled in developmental English to permit correlational analysis of their data. Several provided more than one sample of students; for example, an institution participating in both years of pilot testing, or one testing several groups (such as regular enrollees and students in an enrichment program), could contribute two samples to the analysis. In Table 6 and subsequent tables each institution is identified by a letter and distinct samples from that institution are presented as separate entries.

Over all samples, the median correlation of Reading Comprehension with grades was .32, while that of Sentence Skills with grades was .405. These median values are similar to those reported for paper and pencil tests used for placement. For example, for a sample of four colleges reporting relations of the Descriptive Tests of Language Skills to grades in developmental writing courses, the median correlation of the DTLS Reading Comprehension score with grades was .315, while that of the DTLS Sentence Structure score with grades was .285 (*Guide to the Use of the Descriptive Tests of Language Skills*).

College English

Results for college English courses are presented in Table 7. Four samples from three institutions were available for analysis. The correlations varied substantially from sample to sample; overall the median correlation for Reading Comprehension was .30, while that for Sentence Skills was .305. Comparative results for the DTLS tests show a median correlation of .43 between DTLS Reading Comprehension and grades in college-level writing courses, and of .38 between DTLS Sentence Structure and those

grades. For the College Board Test of Standard Written English, the median correlation with Freshman English grade for institutions using the Validity Study Service was .38 (*The College Board Technical Handbook for the Scholastic Aptitude Test and Achievement Tests*); the median correlation in six samples analyzed by Breland (1977) was .28.

Developmental Reading

Table 8 presents correlations for reading courses for samples from six institutions. Again there is substantial variation from sample to sample; overall the median correlation was .29 for the relation of grades to Reading Comprehension, and .35 for that of grades to Sentence Skills.

Developmental Mathematics

In Table 9 correlations for developmental mathematics courses are displayed. Seven samples from six institutions were available for analysis. Most samples showed substantial relations. The median correlation of grades with the Arithmetic score was .52, while that with Elementary Algebra was .525.

These relations are of the same order of magnitude as those found by Bridgeman and Kline (1980) in a study of the validity of the Descriptive Tests of Mathematics Skills (DTMS). The DTMS Arithmetic Skills test given at the beginning of the semester had a median correlation of .52 with end-of-term grades in developmental mathematics courses (21 samples), while the DTMS Elementary Algebra Skills test had a median correlation of .43 with those grades (18 samples).

College Mathematics

College mathematics course results were available for three institutions. The median correlations with grades (Table 10) were .42 for Arithmetic and .31 for Elementary Algebra. Bridgeman and Kline reported a median correlation of .51 for the relation between the DTMS Elementary Algebra Skills test and grades in seven college algebra courses. Further data for comparison are available for 29 colleges that through the Validity Study Service have studied the relationship between the mathematical section of the SAT and grades in a freshman mathematics course. For those institutions, the median correlation was .33.

In the results summarized above, there is often great variability in the magnitude of the correlations from one sample to another. Median results across samples are much more consistent. In general, the computerized test scores showed average correlations in the range .30 to .40 with grades in English, reading, and mathematics courses; the one exception was that of the relations of test scores to developmental mathematics grades, where the median correlations were over .50. These coefficients are similar in magnitude to those found in studies with standardized paper and pencil batteries and are about as high as can reasonably be expected given the variety of extrinsic factors that operate to attenuate the relationships.

It should be noted that the relations described above arise in courses for which placement was not made on the basis of the computerized test. If the test was employed in placement, the relations of test scores with grades would not be expected to be as great, since the range of scores for students placed in a given level of course work would be restricted; moreover, to the degree that students were appropriately placed, the grades received in the course would also be more homogeneous.

In addition to the results for intact samples, relations with grades were examined for sex and ethnic subgroups in each sample in which contrasting groups of at least 20 cases were available.

Ten instances were identified in which males and females could be contrasted: six samples in developmental English courses, three in developmental mathematics, and one in college mathematics. Only one sample yielded significant differences; males enrolled in developmental English and mathematics courses in one institution showed higher correlations of test scores with grades than did females, probably because the computerized test scores for males were substantially more heterogeneous. No overall tendency

for higher correlations for one or the other group was found.

Three instances were identified in which Whites and Nonwhites could be contrasted: two in developmental English courses and one in developmental mathematics. In only one sample were there significant differences; for students enrolled in a developmental English course, the correlation of Reading Comprehension with grades was greater for Nonwhites than for Whites. Again no overall trend favoring one group was found.

These comparisons are based on small samples, and there is little reason for confidence that the subgroups, depending largely on volunteer examinees, are representative ones; thus they are far from definitive. The strongest statement that can be made from this is that there is no evidence for systematic difference in prediction of grades based on ethnic or sex group membership.

Relations of the Computerized Test to Other Placement Tests

The most systematic basis for comparing the computerized test with paper and pencil tests is provided by data collected as part of the 1984 pilot. The majority of students participating in the pilot were asked to complete a short paper and pencil test constructed from the same pool of items as the adaptive test but sharing no questions with it. Two hundred thirty-four took a reading comprehension test, while 88 completed an elementary algebra test. The results, presented in more detail as a technical bulletin in the Coordinator's Notebook, indicated that there were no systematic differences in level of students' performance associated with the mode of testing and that the relations between corresponding tests, when corrected for test reliability, were nearly perfect.

Institutions participating in the 1985 pilot were asked to provide scores on any available placement tests completed by their students. As shown in Table 11, six schools provided eight samples in which relations of Reading Comprehension to other verbal tests could be examined. The correlations ranged from a low of .17 to a high of .72, with a median value of .50. Similar analyses for the remaining test modules indicated median correlations of .55 for Sentence Skills, .54 for Arithmetic, and .42 for Elementary Algebra.

In most instances, the tests used were institution-specific and the sample sizes were too small to support generalizations about the "true" relations on the basis of one or two samples. Therefore no more detailed presentation of the results is provided.

For similar reasons no specific contrasts will be reported between the computerized test and paper and pencil tests given to pilot students in predicting students' grades. Over all samples and all courses, there were 61 instances in which the magnitude of correlations could be contrasted. One institution, for example, provided scores on three locally administered mathematics tests and both computerized mathematics tests for students enrolled in developmental mathematics; six of the 61 possible contrasts (two computerized tests times three paper tests) arise from their data. Altogether, the computerized test scores showed higher correlations with grades than did the paper test scores in 59% of the contrasts made.

Summary

Seventeen two- and four-year institutions participated in pilot testing of the College Board Computerized Placement Test during 1984 and 1985. Data were available for over 2,500 students who took one or more modules of the adaptive test. Students participating in the 1984 pilot completed a questionnaire describing their reactions, and many took a short paper and pencil test for comparison with the computerized test. In both years of pilot activity, students' grades in relevant courses and their scores on other tests used for placement were obtained.

Reactions of students and of faculty and staff were very favorable. Both groups found the computer easy to use; both liked especially the adaptation of the test to a student's level of skills and the capability to obtain immediate score reports.

The most important component of the pilot was the opportunity to evaluate the effectiveness of the computerized test in predicting grades in English, reading, and mathematics. Grades in single courses constitute a very fallible criterion, confounded by a number of factors. Nonetheless, moderate correlations with grades, with median values generally in the range from .30 to .40, were found for all categories of courses examined except for developmental mathematics courses; in the latter, the median correlations were over .50. These coefficients are similar in magnitude to those found in studies with professionally constructed paper and pencil test batteries.

Table 1
Institutions Participating in Pilot Testing

Institution	Location	Type	Governance
Anchorage Community College	Alaska	Two-Year	Public
Bergen Community College	New Jersey	Two-Year	Public
Central Piedmont Community College	N. Carolina	Two-Year	Public
Delaware Tech. & Community College	Delaware	Two-Year	Public
Evergreen Valley College	California	Two-Year	Public
George Mason University	Virginia	Four-Year	Private
Illinois Central College	Illinois	Two-Year	Public
J. Sargent Reynolds Community College	Virginia	Two-Year	Public
Kirkwood Community College	Iowa	Two-Year	Public
Lethbridge Community College	Canada	Two-Year	Public
Mansfield University	Pennsylvania	Four-Year	Public
New York University	New York	Four-Year	Public
Scottsdale Community College	Arizona	Two-Year	Public
St. Louis Community College	Missouri	Two-Year	Public
Yuba College	California	Two-Year	Public
University of New Mexico	New Mexico	Four-Year	Public
University of Virginia	Virginia	Four-Year	Public

Table 2
Student Reactions to the Computerized Test

How difficult was the verbal part of the computerized test for you? (that is, Reading Comprehension and Sentence Skills)

1%	much too difficult
23%	somewhat too difficult
70%	just about right
5%	somewhat too easy
1%	much too easy

How difficult was the mathematics part of the computerized test for you? (that is, Arithmetic and Elementary Algebra)

9%	much too difficult
37%	somewhat too difficult
41%	just about right
12%	somewhat too easy
1%	much too easy

Did you feel challenged to do as well as you could on the test?

6%	not at all
52%	somewhat
42%	very much

In comparison with a paper and pencil test of the same length, did you find the computerized test

48%	more challenging than a paper and pencil test
39%	about the same
13%	less challenging than a paper and pencil test

Were you nervous while taking the test?

61%	not at all
34%	somewhat
5%	very much

Did nervousness while taking the test prevent you from doing your best?

3%	yes, definitely
17%	yes, somewhat
44%	probably not
36%	definitely not

Which would make you more nervous, a paper and pencil test or a computer-administered test?

43%	a paper and pencil test
11%	a computer-administered test
46%	they'd be about the same

Did you feel any pressure to complete the test quickly?

3%	yes, a great deal of pressure
38%	yes, some pressure
59%	no

How easy was it to use the computer to take the test?

79%	very easy
19%	reasonably easy
2%	somewhat difficult
0	very difficult

Were the directions given by the computer clear and easy to follow?

87%	yes, very clear
12%	reasonably clear
1%	no, not very clear

Did you have any difficulty reading the material presented on the computer screen?

23%	yes
77%	no

Table 3
Mean Scores on the Computerized Test

		Reading Comprehension	Sentence Skills	Arithmetic	Elementary Algebra
All Students	M	75.86	82.95	67.21	55.47
	Percentile	46	33	42	42
	N	2021	1934	1686	1547
High	M	90.89	102.10	90.52	98.59
	Percentile	68	67	72	84
	N	70	62	62	62
Low	M	49.48	55.55	54.34	34.41
	Percentile	19	11	29	20
	N	76	76	125	42

Table 4
Median Correlations Among Scores on the Computerized Test

	Sentence Skills	Arithmetic	Elementary Algebra
Reading Comprehension	.54	.49	.34
Sentence Skills		.44	.36
Arithmetic			.66

Table 5
Coding of Grades

Grade	Numerical Code
A+, A	9
A-	8
B+	7
B	6
B-	5
C+	4
C	3
C-	2
D+, D, D-	1
F	0
WITHDRAWN	–

Table 6
Correlations of Computerized Test Scores with End-of -Term Grades in Developmental English Courses

		Reading Comprehension		Sentence Skills	
		N	r	N	r
Institution	Type				
A	Two-Year Public	68	.40	68	.52
B	Two-Year Public	48	.64	47	.49
B	Two-Year Public	23	.32	–	–
C	Two Year Public	44	.39	44	.16
C	Two-Year Public	62	.25	62	.26
D	Two-Year Public	66	-.05	66	.10
E	Two-Year Public	39	.09	39	.39
F	Four-Year Public	105	.39	83	.42
G	Four-Year Public	21	-.06	21	.49
G	Four-Year Public	26	.05	26	.14
H	Four-Year Public	63	.52	62	.60
		–	–	–	–
		Low	-.06		.10
		High	.64		.60
		Median	.32		.405

Note: N is the number of cases on which the correlation is based;
r is the correlation coefficient.

Table 7
Correlations of Computerized Test Scores with End-of- Term Grades in College English Courses

		Reading Comprehension		Sentence Skills	
		N	r	N	r
Institutions	Type				
A	Two-Year Public	41	.53	41	.47
B	Four-Year Public	47	.31	47	.47
B	Four-Year Public	29	.14	29	.12
C	Four-Year Private	32	.29	32	.14
		Low	.14		.12
		High	.53		.47
		Median	.30		.305

Note: N is the number of cases on which the correlation is based;
r is the correlation coefficient.

Table 8
Correlations of Computerized Test Scores with End-of-Term Grades in College Reading Courses

		Reading Comprehension		Sentence Skills	
		N	r	N	r
Institutions	Type				
A	Two-Year Public	33	.74	33	.69
B	Two- Year Public	38	.41	38	.44
C	Two-Year Public	65	-.11	65	-.01
D	Two-Year Private	43	.03	–	.–
E	Two-Year Public	43	.25	42	.25
F	Four-Year Public	43	.33	43	.35
		Low	-.11		-.01
		High	.74		.69
		Median	.29		.35

Note: N is the number of cases on which the correlation is based;
r is the correlation coefficient.

Table 9
Correlations of Computerized Test Scores with End-of -Term Grades in Developmental Mathematics Courses

		Arithmetic		Elementary Algebra	
		N	r	N	r
Institutions	Type				
A	Two-Year Public	39	.64	38	.64
B	Two-Year Public	52	.52	52	.58
B	Two-Year Public	33	.35	--	--
C	Two-Year Public	34	.20	34	.01
D	Four-Year Public	28	.54	28	.56
E	Four-Year Public	63	.53	63	.41
F	Four-Year Public	61	.44	61	.49
		Low	20		.01
		High	.64		.64
		Median	.52		.525

Note: N is the number of cases on which the correlation is based;
r is the correlation coefficient.

Table 10
Correlations of Computerized Test Scores with End-of -Term Grades in College Mathematics Courses

		Arithmetic		Elementary Algebra	
		N	r	N	r
Institutions	Type				
A	Two-Year Public	49	.41	48	.19
B	Two-Year Public	39	.57	39	.38
C	Four-Year Public	31	.42	31	.31
		Low	.41		.19
		High	.57		.38
		Median	.42		.31

Note: N is the number of cases on which the correlation is based;
r is the correlation coefficient.

Table 11
Correlations of Computerized Test Scores with Other Tests Given for Placement

	Number of Institutions	Number of Samples	Low r	High r	Median r
Reading Comprehension	6	8	.17	.72	.50
Sentence Skills	5	7	.31	.78	.55
Arithmetic	4	6	.09	.76	.54
Elementary Algebra	4	6	.04	.77	.42

Reprinted from *Coordinator's Notebook for the Computerized Placement Tests*, July 1986.

Test-Retest Results

Students retested on the adaptive test of basic skills after an interval of several weeks showed no significant changes in their level of performance. These results suggested that the adaptive test can be used to measure a student's progress within a developmental course or for end-of-course evaluation without concern that practice effects would affect the interpretation of the results.

Forty college students took part in a study of test-retest effects on the adaptive test. The purpose of this data collection was to determine whether students' scores showed practice effects over a short period of time during which no change in their skills would be expected. Changes in test performance could arise because of familiarity with the process of computerized testing, because of familiarity with the kinds of questions included in the test, or because of exposure to specific test questions. If no practice effects were found, two further applications for the test could be considered: to measure progress within a developmental course, or to determine at the end of a course whether students had acquired the skills needed for higher level work.

Participants in the retest were students who had completed all four test modules in an initial testing and who had obtained algebra scores above the chance level – that is, greater than 21 on the Total Right score scale. Students were paid for their participation. The number of days between successive test administrations ranged from 14 to 31 with a median of 21.

No substantial practice effects were found. For the four test modules, the mean change from original testing to retesting was between 1.0 and 4.5 points on the Total Right score scale, which ranges from chance to a maximum of 120 for each module. These changes amount to between .05 and .19 standard deviations relative to results of the initial testing. None of the increases approached statistical significance.

The medians of change scores for the individuals ranged from 0 to 3 points for the four modules. Seventy-six percent of all retest scores fell within 10 points of the original scores, about as expected given no practice effects and typical standard errors of measurement somewhat less than 10 points for students at the ability levels represented by this sample.

Test-retest correlations ranged from .83 for sentence skills to .96 for elementary algebra. Not much credence should be given to differences among the coefficients in a sample of this size; the median of these values, .90, matches closely the reliabilities previously derived from simulation studies.

These results provide an indication that the adaptive test can be used for in-course progress testing or end-of-course evaluation without concern that substantial practice effects will affect the interpretation of the results.

Test-Retest Results

	Reading Comprehension	Sentence Skills	Arithmetic	Elementary Algebra
Initial Scores				
Mean	78.8	83.2	69.8	57.2
S.D.	22.0	23.8	26.6	30.8
N	40	40	39	39
Retest Scores				
Mean	79.9	87.7	71.7	59.1
S.D.	21.9	20.2	28.6	31.8
N	40	40	39	39
t (Correlated Means)	0.11	0.34	0.16	.23
Correlation	.90	.83	.91	.96

Reprinted from the *Coordinator's Notebook for the Computerized Placement Tests*, July 1986.

Computer Adaptive Assessment Testing at Central Piedmont Community College

by Nan A. Webb, Mel Gay, David A. Rhoden, and John Tripp

Background

Early Interest

Attracted by the advantages offered by computerized adaptive testing, Central Piedmont Community College began exploring its feasibility for student assessment and placement in entry-level courses, as early as 1982. In 1983, the college began discussions with the College Board and Educational Testing Service, and by spring, 1984, Central Piedmont Community college became one of three pilot-test sites for the College Board's newly developed instruments, Computerized Placement Tests.

From this pilot implementation of computerized adaptive testing has grown an expanding commitment to the use of the tool. Central Piedmont Community College now has four years of experience in pilot testing, application of the technology to new areas of assessment and service, and research on the test's ability to predict student success in a variety of academic areas. Over 26,000 CPT's have been administered, and the prospects are that the college will seek to implement computerized adaptive testing in programs and services throughout the college.

The College

Central Piedmont Community College is the largest community college in North Carolina, enrolling over 17,500 credit students per quarter. In addition to its main campus in uptown Charlotte, CPCC has four learning centers in other areas of the county and uses more than 250 local sites for teaching various classes. A combination of two-year degree, diploma, and certificate programs is offered to cover the broad spectrum of student and community needs.

CPCC students come from every walk of life and have diverse goals. The college transfer program accounts for about ten percent of the total enrollment, technical students for thirty-one percent, general education for fourteen percent, vocational for five percent, and non-credit extension students account for forty percent. The average age of students at CPCC is thirty-four. The student body is twenty-three percent non-white. Females outnumber males fifty-seven to forty-three percent. The college also has about 1,200 international students representing 102 foreign countries.

Various handicapped students, such as hearing impaired, sight impaired, physically impaired, learning disabled, and multiskills training students, number 250. There are also several social service programs which operate under the CPCC umbrella, including Job Training and Partnership Act (JTPA), Human Resource Development (HRD), and Urban League. Students with handicapping conditions and those in special programs often require special testing services. For instance, some programs require group assessment which may include academic aptitude, vocational, and interest instruments. These programs may test fifty to one hundred people in a group situation during a six-hour day.

The Testing Center

CPCC operated a fully functioning Testing Center when it applied to be a pilot-test site for Computerized Placement Tests. The Testing Center formally opened in 1981 for the purpose of administering course, placement, diagnostic, and certification exams. Staff included a part-time supervisor/counselor, a full-time counselor and

clerk, and several part-time monitor/clerical personnel. Classroom tests used in courses supplied by instructors were administered and monitored continuously from 7:30 A.M. to 8:00 P.M. on regular college-days. Students could drop-in to take tests at their convenience and all class time could be spent for educational purposes.

The Testing Center also provided support for faculty, including test bank operations that generated random tests or multiple forms from a department question bank housed on the mainframe computer. These tests were scanned, analyses run, and returned to instructors. Additional support included help with test development and computers for test authoring or software examination.

Placement tests were administered in group sessions at appointed times. These tests were scored by an optical mark reader and scores were transferred to permanent computer master files. Certification exams, such as Federal Aviation Administration, health care personnel, and state insurance exams, were administered either individually or in groups as scheduled. There were about forty different tests subject to annual administration, and some were administered as frequently as twice a month. Because the Testing Center was in use for various types of tests, students were accustomed to the site and procedures. With the physical facility, service, and personnel already providing the college with many forms of testing, plans were made for necessary adjustment to accommodate CPTs.

Five IBM personal computers and Epson printers, which were purchased through county bond funds in early 1984, were added to the Testing Center to initiate the pilot test. With equipment and personnel in place, computerized adaptive testing was instituted as an added function of the Testing Center.

Pilot Testing

Central Piedmont Community College began operating as a pilot-test site for computerized adaptive testing in May, 1984. Several areas of the college were involved in planning the pilot testing. Support services and instructional departments designed the process and scheduling so that students could participate without confusion.

When the pilot study was announced, 180 students signed up to participate, and 173 completed testing. An incentive of fifteen dollars per student was offered for taking the two-hour test; each student took all four parts: sentence skills, reading, arithmetic, and algebra. ETS personnel participated in the administration of CPTs over a two-week period and subsequently interviewed students, faculty and staff about their impressions and concerns regarding the testing tool. A randomized retest, in June, 1984, was administered to 40 of the 173 to study practice effects. No substantial practice effects were found. Test-retest correlations ranged from .83 on sentence skills to .96 on algebra.

Results of a survey conducted during the pilot test indicated that seventy-four percent of students preferred the computer test to conventional paper and pencil tests; sixteen percent found no difference; and ten percent preferred the traditional paper and pencil tests. Students indicated they found the computer tests interesting, appreciated working at their own pace, and felt comfortable using computers.

Educational Testing Service conducted two sets of analyses on the pilot test. One included all students with available data, and the other was limited to those who took college paper and pencil placement tests during fall, 1983, or winter, 1984. The latter group was thought to provide more accurate comparison with adaptive test scores because there was less time between the two testings.

Correlations were greater when analysis was limited to those with more recent placement test scores. Correlation between the computerized tests and paper and pencil tests were higher with corresponding tests, such as the CPTs sentence skills subtest and the traditional communications test (.58), and the CPTs reading subtest with the traditional reading comprehension test (.63). The data suggested true relations between the adaptive test modules and the closest corresponding local placement test.

ETS expanded pilot testing to include seventeen institutions and to generate a broad base for data analyses. Two and four-year public and private institutions were included. Over 2,500 students

were in this study. Mean scores with percentiles from the summary of the results of the expanded pilot are contained in Table 3.1.

Further Implementation

Next Steps

After completing the pilot test in 1984, key college staff met to evaluate the future of computerized adaptive testing at CPCC. Administrators recognized the convenience on-demand testing could provide. Students could "drop-in" at any time, take the placement test, and complete registration without undue inconvenience of missing time from work or having to make special transportation arrangements in order to meet scheduled testing times. Faculty appreciated the advantage of each student having an individualized test. Since the test adapted to the difficulty level of each student as testing proceeded, the student was not overwhelmed by very difficult questions, or bored by repetitive easy questions. Because of the individualized nature of the tests, the test retest validity was high and allowed retesting without a waiting period.

Testing Center staff determined CPTs could be administered as an added function to the regular routine of testing. Space was available, record keeping for placement tests was already in progress, test security procedures were in use, and several staff members had varying amounts of experience with computers.

Recognizing the advantages of CPTs and wanting to offer these advantages to students, the college examined the feasibility of using both CPTs and traditional placement testing to give students, counselors, and faculty a choice to meet their particular needs.

Necessary Modifications

In order to run a dual system, some means of equating scores from similar tests was necessary. Also, CPCC used a one-page Counselor Interview Sheet that had been designed to provide placement test scores, course recommendations and codes, student ID numbers and other student information, placement results, and counselor suggestions. The student report, generated by CPTs, did not supply this individual information.

Representation from CPCC and ETS discussed the possibility of using CPTs as regular placement instruments if scores could be equated and if a similar Counselor Interview Sheet could be produced at the completion of testing. An agreement was reached that ETS would provide a formula for equating the existing paper and pencil scores to the respective computer test scores, and would program CPTs software to produce a Counselor Interview Sheet when testing was complete. With these hurdles out of the way, plans progressed quickly to begin use of CPTs as a regular placement instrument in early 1985.

Cost was a complicating factor. Five IBM computers and printers had been purchased for administering the pilot tests. Additional ones might be needed before implementing CPTs for routine placement testing. Computer paper would be an additional cost, but no additional Testing Center personnel would be required. The convenience of drop-in testing, immediate scoring and individualized tests had value. Also, at this point, the software was experimental and there was no charge for using it since the data generated would be used by ETS for research purposes. While CPCC would use CPTs as regular placement tools, the data would be a part of the pilot testing which ETS would continue throughout 1985.

At the time of pilot testing, CPCC used an adaptation of Prentice-Hall Communications Test for Writers; Nelson Denny (forms E,F) for reading; and in-house developed arithmetic and mathematics tests. Supplemental tests were used for the nursing and allied health programs. The Counselor Interview Sheet developed by CPCC was also used in the assessment process.

Before dual system testing began in 1985, ETS supplied CPCC with formulas so that scores on computer adaptive tests could be equated to local placement tests already in use. With equated scores, counselors could use either test to place students in appropriate classes. This process allowed CPCC to use Computerized Placement Tests as an actual placement tool. Equated scores were compiled and distributed to admissions counselors, department heads, and faculty in-

Table 3.1 Mean Scores for ETS Pilot Test of CPTs at 17 Colleges and Universities, 1985

		Reading Comprehension	Sentence Skills	Arithmetic	Algebra
All Students	Mean	75.86	82.95	67.21	5.47
	Percentile	46	33	42	42
	Number	2021	1934	1686	1547
High	Mean	90.89	102.10	90.52	98.59
	Percentile	68	67	72	84
	Number	70	62	62	62
Low	Mean	49.48	55.55	54.34	34.41
	Percentile	19	11	29	20
	Number	76	76	125	42

volved in registration in preparation for using CPTs as viable placement tools. Based on equated scores, cut-scores were established for CPTs. ETS also modified its testing software to provide a report comparable to the locally developed Counselor Interview Sheet.

Staff Training and Evaluation

Before implementation began, preparations were made to orient students, staff, faculty, and administrators to the new form of testing. In preliminary meetings with counselors, expected problems such as new procedures for record keeping, scheduling students, and transfer of test data were identified, and plans were developed for explaining the system to college personnel. A series of workshops was planned to give college personnel hands-on experience with the tests, as well as detailed instruction in their use.

Workshops were designed for eight people, and were scheduled in various time frames. In such small numbers, each person could use a PC, feel at ease to ask questions, and personally take the CPTs. Information about the tests was distributed with sign-up sheets, and administrators were informed of progress. Strong administrative support, including department heads' scheduling their entire staffs to attend the workshops and public comments by key leaders, assured success of the workshops.

These workshops were held in the Testing Center where placement and course tests were given and where CPTs would later be administered. Participants were given a background report on the development of the tests, and their advantages were discussed. The fact that a student could work at his or her own pace was a great advantage, and was demonstrated in the workshops as faculty and staff worked at their own pace at computer terminals. Knowledge of computers was not necessary for testing, and faculty and staff proved this point as some of them had no previous experience with computers.

Participants anticipated ways the students would approach the testing environment, computers, printers, and reports. They were encouraged to ask questions, make comments, or express concerns about the tests, reports, and procedures being considered. These comments allowed some immediate changes and posed ideas to be considered in the future. A few specific instructions for students were identified, and these were put on cards placed beside the computers. The workshop sessions allowed both faculty and Testing Center personnel first-hand experience with the new tests, and provided opportunity to determine how each could help students in an actual testing situation.

A hand-out designed to explain computer testing to students was given to personnel so they could give feedback on clarity and inclusiveness. In order to check for procedure realism, workshop participants were instructed in the same manner planned for later use with students. However, instructors were told to include wrong answers so they could observe how the computer would regress to easier questions when a student gave an incorrect answer. Workshop participants sampled all tests (sentence skills, reading, arithmetic, and algebra), but completed only one to save time. One test must be completed to generate an individual Counselor Interview Sheet. Each participant received a report indicating total score and percentile.

From the orientation workshops, functional problems such as lighting, glare, and computer height were solved before students actually took the tests. The instructors pointed out a few minor ambiguities that should be rectified so students would not be confused. Instructors and staff viewed the test from their particular perspective and offered constructive comments concerning scoring, scheduling, and test content. They particularly liked the possibilities of administering the tests on a drop-in or individually scheduled appointment basis rather than assigning students to take placement tests in scheduled group situations at a future date.

With the "bugs" worked out of the system, computerized adaptive testing was implemented as an alternative to traditional placement tests. Information about the new alternative was distributed broadly to students and staff throughout the college.

Scheduling and Personnel

A student seeking assessment begins in the admissions office, which has responsibility for scheduling on-demand testing. The admission office verifies the availability of a testing station by a call to the Testing Center. Unless no station is available, the student is referred to the Testing Center, completes assessment testing, and continues the advisement and registration processes. Scheduling begins with eight slots at 8:00 A.M.; three slots at 9:00 A.M. This pattern is repeated until 6:00 P.M. when the last eight are scheduled. By this method, computers are fully utilized and students have up to two hours to complete testing. Students taking three or four tests may use the two hours, but students taking only sentence skills or a combination of two, finish in much less time.

Placement test scheduling continues through final registration. On these days, the time slots are usually full, but an occasional drop-in can squeeze in when someone finishes early. Retest applicants follow the same procedure and come through the admissions center for scheduling. Retests are not given without faculty, counseling or administrative authorization. However, retest may be given without a waiting period because the student will not get the same questions from the CPTs question bank.

At CPCC, personnel already involved in testing have been trained to administer Computerized Placement Tests. While one full-time person takes primary responsibility for booting, running rosters, and keeping up with disks in use, all members of the Testing Center staff assist with starting students or generating individual printouts.

Record Keeping

Each student arriving at the Testing Center to take the CPTs signs a register which includes name, social security number, program code and test(s) to be taken. The testing assistant adds the disk serial number, indicates tests taken, and approximate time testing began. The disk number is very important if future verification is needed. This number will immediately identify where original records are located and in case of lost information, makes retrieval much simpler.

The Interview Sheet or individual report, which is printed at completion of testing, is delivered to the counseling department at the end of each day. The counseling department notifies the student of an appointment time for counseling. These reports are printed on four-part paper so this report may be given to counseling and permanent records, veterans, or financial aid departments, as well as the student. On final registration days or when time is limited, the student may receive a copy of the report and go directly to registration to complete the enrollment process. It is expected that the student will receive program counseling as soon as possible.

Through a computer program, the scores on the disk are uploaded to the mainframe once or twice daily, depending on registration needs. This is usually done from the Testing Center in late afternoon. These scores are transferred to the mainframe computer records at night and are available by terminal the next day. Uploading provides accuracy not available by data entry, and it conserves personnel time. The process requires conscientious regularity on the part of Testing Center personnel and Computer Center personnel to ensure this process is accomplished on a daily basis.

Testing Center personnel keep records of CPTs, including daily, weekly, and complete historical rosters. These provide hard copy accessible by date. In addition to these, each student's permanent record contains a hard copy of the Counselor Interview Sheet and scores maintained on the mainframe student database which are accessible by social security number.

Using Placement Scores

Using the CPTs scores was initially facilitated by equating these scores to traditional paper and pencil tests already in use. At CPCC, all degree, and most diploma and certificate students, are required to take placement tests. Registration for college-level English or math classes requires appropriate placement tests, regardless of the planned program. Students taking special interest classes are not required to meet these standards.

Developmental classes in English, reading, and mathematics are offered for students who are deficient in these skills. In 1985, cut-scores for these classes varied according to college program of study. However, the Advancement Studies Department, the college's developmental program, revised cut-scores for reading and arithmetic to assure uniformity in skills levels for all developmental courses. Currently, all programs use the same cut-scores for reading and arithmetic.

Using the 50th percentile scores based on a national sample of CPTs and 50th percentile scores based on four quarters of CPCC student scores, as well as other data, the Advancement Studies Department established cut-scores for three reading levels and two arithmetic levels below college level. These courses provide opportunities for students to complete successfully these necessary skills in attainable segments.

At CPCC, courses below college level are numbered in the 9000 series. Reading course 9505 is a developmental course that emphasizes fundamental skills, including basic vocabulary, comprehension, and study skills. Course 9510 moves into finding main ideas, drawing inferences from these ideas, and evaluating ideas in preparation for reading college textbooks. Reading 9312 is a speed-reading class designed to increase reading rate while maintaining a high level of comprehension. Cut-scores have been set for each of these developmental classes.

MAT 9500 is designed to teach arithmetic operations and problem solving relating to whole numbers, fractions, decimals, percent, and measurement. MAT 9501 is designed to teach applications of mathematical functions, including ratio and proportion, percent, metrics, geometry, graphs, and introduction to algebra. Specific cut-scores have been set for each of these classes so students may gain a mathematical foundation in MAT 9500 before advancing to MAT 9501.

Students are not admitted to college-level English or math classes until prerequisites are met, or placement scores indicate the student is prepared to be successful at the college level. By using cut-scores to place students at the appropriate level of study, students may progress through needed skills development and their opportunities to succeed are greatly increased.

Evaluating CPTs Implementation

CPTs have been in use as a regular placement instrument at Central Piedmont Community College since February, 1985. At the same time, traditional paper and pencil test have been offered about three or four times monthly because some counselors want diagnostic information, some social service agencies want group administration, and anyone preferring this method has that option. Also, tests are mailed to high school counselors for administration when applicants are

distant from campus or when other conditions require paper and pencil testing.

Perceived Benefits

Administrative Personnel recognize the benefit to students and college that on-demand placement testing provides. The convenience of being able to test on-demand or at a time suitable for the student may be extremely important to working people, single parents, or homemakers. As community colleges serve increasing numbers of part-time, non-resident adults, the benefit of on-demand placement testing becomes extremely important to many types of educational institutions.

Testing and immediate results allow students to complete preliminary work on final registration days without delaying college entrance. Accurate placement has been shown to lower attrition. Traditionally, many students who inquire about college do not follow through with plans. Often students agree to come back to the school in two or more weeks to take placement tests as scheduled, but a large percentage of these students do not arrive for testing. In the weeks of waiting for the placement tests, interest waned and distractions predominated. Records kept by CPCC admissions for the first five months of testing from February to May, 1985 indicate that an average of fifty-eight percent of students completed traditional placement tests as compared to seventy-eight percent completing CPTs. Being able to test and complete additional registration requirements at the time a student is initially interested seemed to sustain student commitment.

Four Learning Centers, located off campus, use CPTs and can complete testing and registration for students in their own locale. The same orientation procedure was used to acquaint Learning Center personnel with computer tests before they were begun. The Testing Center on campus coordinates procedures and transfer of scores to the mainframe so records are uniform and current.

Handicapped students also appear to have benefitted from computerized testing. Students with handicaps involving loss of motor skills appear to find touching computer keys much easier than managing a pencil. Also, untimed tests are usually more convenient for the handicapped student.

Expressed Concerns

Some faculty and counselors were concerned that the question range did not include questions suitable for students with skills well below college-level (below ninth grade). This may pose a problem for the test was designed for college placement and those far below this level may need another test. Some counselors felt that the single score in arithmetic and algebra did not give them the diagnostic information they wanted. Since the test was not designed as a diagnostic tool, counselors needing diagnostic information may need another test.

Some counselors reported finding individuals who appeared to have higher scores on sentence skills than on reading subtests and questioned why students who could perform well on one did not perform similarly on the other. ETS reported reading questions required more analysis and synthesis while questions in the sentence skills subtest could be answered with lower-level skills. Counselors working with college transfer students did not report this difference. The better-prepared students appear to have better analysis and synthesis skills. Local research is in progress to determine if communications cut-scores should be raised both on CPTs and paper and pencil tests.

The math department, through tracking research, has determined that CPTs and the in-house math placement test are equally effective in placing students into college-level courses. Their research compared student grades in college-level and the highest-level developmental math course with placement test scores. This comparison indicated cut-scores on both pencil and paper test and CPTs were reasonable. It was recommended that students with scores very near the cut-score should be advised to take the developmental course. However, about half of these students would be successful in the college-level course. Success in the developmental course virtually assured success in the college-level course. The Computerized Placement Tests were designed to assess entry-level college skills, and it appears the tests accomplish this function with consistency.

Table 3.2 Test Scores of CPCC Students Taking Computerized Placement Tests, 1985-1987

	Reading Comprehension	Sentence Skills	Arithmetic	Elementary Algebra
Number of Cases	7,714	8,595	4,481	5,425
Mean Score	75	84	57	48
Median Score (50 percentile)	77	87	53	38
Range[1]	3 -120	10 -120	19 -120	9 -120

[1] 120 is a perfect score

Ongoing Research

CPCC has administered 26,215 individual Computerized Placement Tests over a three-year period. Students may take from one to four tests depending on their program. For example, college transfer students would take sentence skills, reading, and algebra while students going into fire protection technology or law enforcement take only sentence skills. General education degree students take sentence skills and reading, and medical records students take sentence skills, reading, and arithmetic. While arithmetic and algebra are not required for the same program, students, may take both tests if they feel unprepared in algebra or had it years ago.

The composite results of this testing demonstrate the broad range of skills of students entering CPCC. Table 3.2 details the mean, median, and range of scores earned by students taking the CPTs option for entry-level assessment. These provide useful data on CPCC students, including a baseline with which to compare future CPCC students.

However, using the CPTs also provides the advantage of comparison to normative data being developed by ETS nationally at both two and four-year colleges. For instance, comparison of CPCC scores with those of entering students at five four-year colleges included in an ETS pilot, demonstrates the differences in the range of skills of students enrolled by these different types of institutions. As the data base is developed further, comparisons with comparable community colleges in the state and across the country can provide additional insight into relative student preparedness.

Future Plans

The top priority of the Testing Center is to be as automated as possible in order to offer fast, efficient assessment service to students, faculty, and counselors. This goal is complemented by its second goal to use the most up-to-date equipment, test procedures, and recently developed standardized tests. These two goals demonstrate the high priority placed on service to students and faculty in the most efficient and modern manner.

In the four quarter period, summer 1986 through spring 1987, the Testing Center administered 84,394 tests of various kinds. This figure almost doubled the 45,844 tests administered from April 1981 through March 1982, which was the first year of the Testing Center Operations, nearly one hundred percent growth in five years. During the fall quarter, 1987, a daily average of 415 students used the center for course placement and certification testing. CPTs ranged from one or two per day to about thirty-five per day.

Testing can be a stressful situation for students, faculty, and staff personnel. Understandably, students are concerned that they score as high as possible; faculty are concerned with fairness, test

security and efficiency. Testing Center personnel share these concerns and their professional concerns for accuracy in administration and interpretation of tests. Jobs done by machines reduce the possibility of inevitable human error. For this reason, automation on the premises becomes highly desirable. The process of computer testing and automatically transferring scores to records facilitates accuracy and speed in keeping placement test scores current for administrative use.

Today, seventy-five percent of placement tests at CPCC are given on computers. Networking for placement tests appears to be a probability and may prevent some recurring disk problems by eliminating dust and handling of disks. Networking would allow one or more placement tests to be administered to a student as needed. Currently, disks are configured to have a combination of tests on one disk. When using a combination disk, some tests may be skipped when not required for that particular student. Skipping tests is not efficient because there is a charge for each test supplied, whether used or not. With networking of the tests, it may be possible to key in the needed configuration as the student begins and, hopefully, no further instructions or assistance would be needed.

The mainframe computer might become the host for placement tests. In this case, a student taking the test could have scores transferred to the appropriate department with no delay. Reports could be generated in any form programmed, and a site license agreement might allow continuous assessment from various locations.

Bar code scanning that interfaces with the overall electronic communications can reduce the time it takes each student to sign in, produce identification, and be issued the correct assessment. With bar code scanning, the student could insert identification that would initiate a record of name, social security number, tracking of timed tests, and test document number or disk number. When the student exits, the same identification check could be used to ensure student exited as prescribed. Regardless of heavy volume, test security and student identification are prime concerns in protection of test materials. To be effective, bar code scanning needs to mesh with computer records of students, courses, and appropriate lists of qualifications.

Data retrieval may require several methods in order to meet all needs. If placement test scores could be retrieved by name, social security number, or other appropriate source, it would be convenient to check needed information. Because most records are kept by ID number or social security number, ability to access them by this method is essential.

Computerized testing has expanded exponentially the possible applications of testing. Highly efficient, accurate, and individualized tests provide opportunities to expand services to new populations. Providing assessment services to the public at large has potential as a recruitment tool. Self-testing, immediate scoring, on-demand testing, touch screen tests for manual dexterity, and hands-on applications in technology or other specialty areas only begin the list of possible innovations.

Reprinted with permission from *Computerized Adaptive Testing: The State of the Art in Assessment at Three Community Colleges;* July 1988.

Nan A. Webb is Counselor, Test Development and Research; Mel Gay is Dean of Student Development; David A. Rhoden is Testing Center Supervisor; and John Tripp is Director, Testing Services and Veteran Affairs at Central Piedmont Community college in Charlotte, North Carolina.

Uses of Computerized Adaptive Testing at Santa Fe Community College

by G. Thomas Delaino, Robert K. Myers, and Pat Smittle

Introduction

College and State

Santa Fe Community College, located in Gainesville, Florida, serves a two-county district in North Central Florida. It is a rural area of approximately 1,200 square miles with a population of 207,000. Gainesville is the largest population center with 95,000 nearly half of whom are college students attending the University of Florida (35,000) and Santa Fe Community College (9,000). The college operates as part of a state system of twenty-eight community colleges.

Since it was established in 1965, Santa Fe Community College has espoused a commitment to help all students reach their full potentials. In 1981, this commitment led the college to adopt a required assessment and placement policy. This college policy was strengthened in 1985 when the State Board of Education adopted a rule that requires all first-time-college students who are seeking degrees to be tested prior to their initial registration. Students who indicate skill deficiencies on one of four state-authorized tests (SAT, ACT, MAPS, and ASSET) are placed in developmental writing, reading, and/or mathematics courses. Students are required to demonstrate competency in the deficient skill areas prior to enrolling in college-level coursework.

State guidelines hold the institution accountable for placement and monitoring of developmental students. The state rule requires that all student placement transactions are auditable. Therefore, the college's developmental studies advisement staff was faced with the difficult task of ensuring accurate initial placement of students into developmental courses. The staff successfully met challenges to the existing system and closed loopholes which students had previously used to bypass mandatory placement.

Increased Demands

Those state-mandated guidelines caused the full-time equivalent enrollment in developmental courses to increase from three to nine percent of the total college enrollment. Specifically, forty-five to fifty-five percent of the entering freshmen at Santa Fe Community College are currently required to enroll in at least one developmental course. This increase in enrollment, coupled with the demand for accountability, created the need for additional instructional staff, academic advisors, and instructional space. In addition, the admissions and registration process was significantly affected. Prior to the registration process, testing personnel administer assessment test to thousands of incoming students each year. A same-day student registration requires immediate scoring, recording, and reporting of assessment results. Although the college's on-line registration system was updated to monitor student placement in accordance with assessment results, the entire process of scoring and entering data continues to be time consuming and laborious.

Once the students were properly placed, the developmental studies faculty faced the difficult task of helping students develop the skills required for competency in each area. Students cannot exit from developmental courses until they demonstrate competency by passing a departmental examination that addresses state-mandated skills. To assess student progress toward the exit competency level, frequent testing with reliable assessment tools is necessary. How-

ever, given the large number of students enrolled in developmental classes, assessment instruments that are efficient and cost effective in terms of student and staff time needed to be found.

Solutions

In July of 1986, at the national conference of the League for Innovation in the Community College on the use of computers in education, the college staff was introduced to a computerized placement test developed by the Educational Testing Service and the College Board. The unique features of this instrument and testing format addressed a number of the college's concerns with current assessment activities and with the implementation of mandatory state placement requirements.

An initial conversation with representatives from the College Board , International Business Machines, and Miami-Dade Community College led to the formulation of a project sponsored by the League for Innovation. The primary focus was for each college to establish a testing site for studying the feasibility of using the College Board's Computerized Placement Tests (CPTs) as an alternative to the state-mandated paper and pencil tests.

The nature of the state's rule did not allow substitutions for the prescribed assessment instruments. Therefore, a primary goal of the project was to collect testing data from the CPTs and create a research design which would validate the test for entry-level testing and placement. A secondary goal was to evaluate the use of the CPTs to monitor progress of students in developmental classes. During the course of the project, staff would also experiment with the CPTs as tools for supplemental testing and for motivational testing with special clientele.

The Project

The project was divided into two phases that were designed to accomplish defined goals. The goals of Phase One, which commenced in September, 1986, and ended in December, 1986, were to plan, organize, and establish a testing center where the Computerized Placement Tests (CPTs) could be administered. Also during this time, the faculty, staff, and students would be provided the opportunity to become acquainted with the concept of computerized testing. The goals of Phase Two, which began in January, 1987, and ended in August, 1987, were to conduct large-scale testing sessions which would provide statistical data that could be used to substantiate a request to the Florida Commissioner of Education to name the CPTs and an additional test to be used for entry-level placement in the community college system.

The research activities during these two phases were designed to determine the following: the appropriateness of the CPTs skill content, the student response to computerized adaptive testing, the establishment of accurate cutoff scores for efficient student placement, the relationship of the CPTs to other placement tests, and relative costs of the CPTs.

The testing center would also provide the opportunity to address the secondary goal, the evaluation of the potential of computerized adaptive testing for educational purposes other than placement testing. During the course of the project, the use of the CPTs to monitor the progress of and provide feedback to students in developmental classes became very important to the development studies staff.

Phase One: Developing a Model Testing Center

Phase One consisted of two stages: a planning stage and an implementation stage. During the planning stage the resources needed to conduct the project were determined. In the implementation stage, the testing center was put into operation, and the instructional staff was introduced to computer adaptive testing.

Planning

In the planning stage, the first aspect of the operating environment to be addressed was space to house the twenty-five personal computers and printers supplied by IBM. The location had to be in an area where reasonable security measures would prevent damage or loss to the hardware or

the CPTs software furnished by the College Board.

The physical layout of the testing center would have to maximize the use of the space and present an attractive and pleasant surrounding. Also, the space had to be centrally located to allow convenient access. A self-contained room in the Learning Lab area was chosen as the site for the computer testing center. Varied furniture arrangements were studied in an attempt to provide maximum utilization of the space while providing comfortable individual testing stations. Window coverings were purchased to reduce the glare on the computer screens and provisions for adequate electrical power were made.

Computer expertise was not required to operate and maintain the CPTs software, but it was decided that the college commitment to establish a model computer testing center would require additional staff with computer experience. Although extensive training and commitment to underprepared students has led the learning lab/developmental studies staff to produce an outstanding program – one that was selected to be included in the National Directory of Exemplary Programs in Developmental Education – it was acknowledged that the administrative and instructional staff, who were highly trained in traditional instructional methodologies, were practically computer illiterate. It was evident that the same commitment that had led them to search for improved teaching materials and methods in the past would prevail in the use of this new instructional technology, but in order to provide support and guidance during the learning period a computer lab manager was needed.

The computer lab manager would monitor the set-up and use of the hardware, as well as have a working knowledge of the software. In addition the manager would have the responsibility for the daily operation of the testing center, faculty and staff training, and security and maintenance of equipment and materials provided by IBM and the College Board. A graduate of the college's electronics program presented these qualifications and was employed as the computer lab manager. With space designated and furniture and additional electrical power provided, the planning stage of Phase One was completed.

Pilot Implementation

Implementation began with the first delivery of hardware. The agreement with IBM provided that the hardware shipments would be made on an "as available basis." The first delivery arrived in August, 1986; the last delivery in January, 1987. As each unit was received, it was uncrated, diagnostically tested, and installed at a testing station. In order to minimize the noise level of the computer lab, the decision was made to limit the number of printers that would be used to generate test reports. As testing stations became operational, the computer lab staff became familiar with the operation and features of the two software packages by studying the associated documentation, taking the four Computerized Placement Tests, and working through the College Board's Computerized Placement Management Software (CPMS) tutorial. Other faculty and staff members were also invited to become familiar with the CPTs package.

When an adequate number of testing stations were available, an initial group of 194 students was tested to provide a "trial run" of the center. These students were enrolled in the College Preparatory Program where they had been placed according to their test scores on the traditional assessment tests, the ACT and SAT. They were selected because of their instructors' expressed interest in the project. Most of the testing was accomplished during class periods where the time ranged from fifty to seventy-five minutes and testing was limited to the discipline in which the students were enrolled.

Since computerized testing was a new concept at Santa Fe Community College, participating students and faculty members were asked to express their opinions on the use of the computer for test administration. The students completed an evaluation form while the faculty members were informally polled.

The results of the student evaluations showed the student attitude toward computer utilization

was enthusiastic and positive. Ninety-seven percent "liked it a lot" or thought "it was okay." Ninety-three percent were "comfortable using the computer," and sixty-nine percent indicated that they learned something about a computer. Ninety-one percent "had fun with the computer," and fully ninety-four percent wanted more opportunity to work with computers.

The overall result of the informal poll of faculty members was positive, but it was tempered with a "wait and see" attitude. There was little doubt that the students were more comfortable than faculty when using the computer. Phase One, its goals attained, was concluded. At this point, attention turned to the commencement of Phase Two and the achievement of its defined goals.

Phase Two: Research and Documentation

Building upon the experiences in Phase One, the college launched a large scale testing program in January, 1987. In Phase Two approximately 2,600 Computerized Placement Tests were administered for the purpose of examining the test as a tool not only for entry-level testing, but for progressive testing, supplemental testing, and motivational testing for special clientele.

Entry-Level Assessment

Santa Fe Community College defines entry-level testing as the assessment of the basic skills of incoming students which serves as the basis for accurate course placement. The college, having not yet received authorization from the State Department of Education to use the CPTs as an assessment and placement test, attempted to simulate the conditions of entry-level testing by testing students who had already been placed in their courses by their ACT or SAT scores in accordance with the state-mandated rule.

To minimize any advantage that additional instruction might give participating students, the entry-level testing of Phase Two was conducted through the first three weeks of the term during class periods which ranged from fifty to seventy-five minutes. In light of this time constraint and in an effort to use the CPTs feature of an untimed test, the majority of students took only the test that was associated with the course in which they were enrolled. This restriction assured that the students would have ample time to complete their test sessions within the scheduled class period. In the instances where students were able to take the complete battery of four CPTs, they were able to complete their testing sessions in an average time of less than one and one half hours. Of the approximately 830 students who participated in the simulated entry-level testing, 230 were enrolled in college-level English, reading, and/or mathematics courses, and 600 were enrolled in developmental writing, reading, and/or mathematics courses.

Skills Content

Florida State Board of Education Rule 6A-10.033 – College Preparatory Competencies identifies the distinction between college-preparatory and college-level skills. A comparison of the skills addressed in the CPTs with those identified in the rule indicates that the CPTs cover all college preparatory skills with the exception of spelling. This is illustrated in Table 4.1.

Placement Scores

To assure all students the opportunity to profit from the college experience, proper placement in entry-level courses is essential; therefore, careful attention was given to the determination of cut-scores for initial placement into college coursework. Various methods of determining cut-scores were explored. The course structure and curriculum at Santa Fe Community College had been designed to accommodate the placement of students according to state-mandated ACT scores. Therefore, the most important factor in determining cut-scores was to maintain the same student distribution in developmental and college-level courses as had been assigned by the ACT during the previous year. With this factor in mind, the data were analyzed and the scores were recommended for initial placement into developmental courses. Initial observations indicate that reliable cut-scores for placement within levels of developmental courses may be determined by using the CPTs. Continued research is planned.

Table 4.1
Comparison of Florida College Prep Competencies and Skills Measured by Computerized Placement Tests

Florida College Prep Competencies	Skills Addressed By CPTs? Yes	No
Reading/CPTs Reading Comprehension:		
1. Main ideas	X	
2. Supporting details	X	
3. Words in context	X	
4. Author's purpose	X	
5. Author's tone	X	
6. Valid arguments	X	
7. Explicit and implicit relationships within and between sentences	X	
8. Bias	X	
9. Fact/opinion	X	
10. Inferences	X	
11. Conclusions	X	
Writing/CPTs Sentence Skills:		
1. Word choice	X	
2. Sentence structure	X	
3. Grammar	X	
4. Spelling		X
5. Punctuation	X	
Mathematics/CPTs Arithmetic and Algebra:		
1. Arithmetic	X	
2. Introductory Algebra	X	
3. Real numbers and their properties	X	
4. Basic operations with linear expression	X	
5. Factoring of algebraic expressions	X	
6. Solutions of linear equations	X	
7. Inequalities	X	

Relationships with Other Tests

During the pilot testing period of the CPTs, additional data were collected, including CPTs results, student ACT and SAT scores, and student course grades. These data were provided to the Office for Management Information Services at Santa Fe Community College to determine if there was a relationship between the CPTs and presently used placement tests. The analysis concluded that the low correlations found were the result of testing students in developmental classes for which ACT and SAT instruments have questionable utility.

Student Response

Participating students were also given a questionnaire which solicited their impressions of the test. The results of the student questionnaire showed that student had positive attitudes toward the CPTs. They reported that the computer was easy to use, and they preferred a computer test to a paper and pencil test. Table 4.2 illustrates the primarily positive attitudes of the students toward the CPTs.

Table 4.2
Summary of Student Evaluations, Spring 1987

Which would make you more nervous, a paper and pencil test or a computer-administered test?

38%	a paper and pencil test
19%	a computer-administered test
43%	no difference

How easy was it to use the computer to take the test?

62%	very easy
25%	reasonably easy
12%	somewhat difficult
1%	very difficult

In comparison with a paper and pencil test of the same length, did you find the computer test...

8%	more tiring than a paper and pencil test
32%	about the same
60%	less tiring than a paper and pencil test

Which would you prefer to take?

63%	a computer-administered test
15%	a paper and pencil test
22%	do not prefer one over the other

Cost

There are two main areas of cost to be examined when considering the use of the CPTs for initial placement. For software, the cost of an entire battery of CPTs is three dollars. In comparison, the cost of an ACT test is eight dollars. Given the number of assessment tests administered by Santa Fe Community College last year, this represents an approximate savings of $10,000 per year. In hardware, the initial investment of funds may appear to be prohibitive, but this is not necessarily the case. For institutions which are presently using the ACT, the per year savings from the cost of the test, if applied to the hardware investment, would quickly diminish the actual amount of capital outlay. For institutions which would not realize this savings, capital outlay could be minimized because a testing program could be initiated with only a few computers. For both types of institutions, when testing is not being done, the hardware can be used for instructional purposes, or even for administrative applications.

Progressive Testing

During the course of the project, the developmental studies staff had the opportunity to apply and evaluate the CPTs as tools for ongoing assessment. Although this was not the primary goal of the project, it became highly important to the staff because of the challenge of teaching developmental classes.

The college has a long history of commitment to underprepared students and has earned regional and national recognition for this activity. However, the mandated placement of students by state rule intensifies an old problem: low student motivation. When the project began, the developmental studies courses were experiencing attrition and failure rates of approximately forty-five to fifty percent. One of the chief problems identified by staff was lack of motivation caused by a drop in the students' academic self-concept. Their high school experiences had led them to construct an academic self-concept that did not stand up to the inspection of college admissions assessment.

Motivating with Assessment and Feedback

Before this project was even conceived, the staff had searched for instructional strategies which would attack the problem just described. A recent national study group on the *Conditions of Excellence in American Higher Education* identified three conditions critical to maximize student growth and learning: student involvement, high expectations, and assessment and feedback. This third condition, assessment and feedback, was seen as an excellent target for computer applications. The introduction of the CPTs, with its unique testing format, provided the tool necessary to test this one condition and its use in helping students achieve competency in the basic skill requirements.

Providing students with assessment information about their own performance can be an essential ingredient in effective learning and can serve as a powerful lever for involvement. However, this approach requires testing beyond what is needed to evaluate students for grading purposes, and therefore, places even greater demands on already heavy faculty workloads. They were being asked to consider additional assessment testing beyond their need to evaluate student performance for grading purposes.

Given the state-mandated emphasis on skills, it was assumed that if faculty were satisfied with the skills measured by the CPTs, their concerns regarding administration, scoring, recording, and feedback of assessment results would be resolved once they were familiar with the testing software. It was also predicted that students would respond positively to information about their performance in relation to faculty expectations, and, as a consequence, enhance the likelihood of their involvement in the learning process.

Experimenting with Progressive Testing

The capabilities of the CPTs for monitoring student progress, for generating multiple forms of the test, presenting questions within a test-taker's range, and providing immediate feedback, plus the overwhelming student preference for a computer-administered test, were major factors in the

decision to experiment with progressive testing in developmental studies classes. Approximately 670 students, all of whom were enrolled in developmental writing, reading, and/or mathematics classes, participated in the progressive testing project. They were tested three times: at the beginning, at the mid-point, and at the end of the semester. The tests were administered during fifty to seventy-five minute class periods, which allowed ample time to complete the tests. The students took only the specific test which coincided with the course in which they were enrolled. A student was deemed to achieve progress if his or her score on the mid-term or final test was greater than his or her score on the initial test.

For the students who realized progress, the CPTs results helped to increase their confidence level and motivated them to set higher goals. The CPTs provided warning signals for the students who did not demonstrate progress. The test results acted as a springboard from which their instructors could institute student/instructor conferences where individual circumstances could be examined and intervention strategies enacted.

The use of the CPTs as a yardstick to measure student progress appears to be a valuable application in a developmental education course. It provided benchmarks of progress for all students. If a student achieves progress, the test supplies positive reinforcement and bolsters confidence. But more importantly, if a student does not demonstrate progress, the CPTs can provide early information that signals instructors or counselors that a need for intervention has arisen. Though the structured testing stages of Phase Two have drawn to a close, this opportunity for early intervention has influenced developmental reading instructors to continue its use for progressive testing.

Supplemental Testing

Another application of CPTs is supplemental testing. This has been defined at Santa Fe Community College as testing that is administered to students who feel the need to be placed in a developmental course even though their ACT or SAT scores place them in a college-level course. A subsection of the state-mandated rule requires that any student who seeks enrollment in a developmental course must first demonstrate his need through an assessment technique, and CPTs are used as the assessment vehicle.

A rigid schedule was not followed for supplemental testing. These untimed testing sessions were administered by appointment at times which were chosen by the test takers. Without the inherent flexibility of scheduling that CPTs offer, such an arrangement would have been impractical.

The students who sought supplemental testing represented a cross section of the student population. Some had been away from school for a number of years and were looking for accurate course placement. Others had not succeeded in college-level courses and sought admission to developmental courses which would refresh their skills. Still others were seeking confirmation that previously acquired skills had not eroded through disuse.

Motivational Testing for Special Clientele

The concluding CPTs application of Phase Two, motivational testing for special clientele, was administered to high school students who had been selected to participate in the College Achievement Program (CAP). In an effort to recruit and retain minority students, Santa Fe Community College conducts this special summer academic program for minority students who have indicated an interest in attending the college. These high school graduates attend an on-campus, six-week session during which they study writing, reading, mathematics, and college survival skills.

The goal of this CPTs application was to provide motivation to the forty-three participating students. With this in mind, the students were advised prior to the first testing session that the CPTs were self-evaluation tools. These sessions were conducted at two-week intervals during sixty-minute periods and all sections of the test battery were administered.

This application produced an unexpected result. The CAP students were excited at the prospect of using computers. In response to this attitude, the students were introduced to Computer Assisted Instruction (CAI) in the area of mathematics. Their enthusiasm did not ebb. They eagerly attended the CPTs and CAI sessions, and in the infrequent instances when a student was absent, he or she made arrangements to make up the missed session.

Perhaps an observation that was related by one of the test administrators best describes the atmosphere of these sessions: A short, piercing scream claimed the attention of everyone in the room. It had come from a young lady who sat staring at the computer screen in front of her. The test administrator went to her and asked, "What's wrong?" She pointed to the screen and whispered, "Look at my score." The score was 87, and the test administrator told her, "Anyone who scores that high can scream anytime they want." She screamed again!

Fully 91.2 percent of the CAP students demonstrated progress in college preparatory skills measured by the CPTs during the six-week session. The degree of the CPTs contribution to this progress could not be empirically measured, but the potential for using computerized adaptive testing to increase motivation was clearly demonstrated.

Summary Results

At the onset of the project a number of the potential advantages which CPTs had to offer, such as individualized tests and flexibility in test scheduling, were obvious. As the project progressed, less obvious advantages came into focus for students and staff.

For students, CPTs with their designed property of branching, provided challenging but not frustrating testing sessions. Immediate feedback and untimed testing sessions were other characteristics of CPTs that students found favorable. Progressive testing afforded the students the unusual opportunity to examine their skills development in the course of a semester. The results of the student surveys showed that the large majority of students preferred the computer-administered test over a paper and pencil test.

The CPTs and management software furnish the college with an efficient, compact media in which to store the accountability documentation on college preparatory students which the State Board of Education requires for entry-level and supplemental testing.

For staff members, the CPTs furnished a readily accessible and easily maintained method of monitoring the progress of developmental students. Reading instructors adopted the adaptive reading test as an alternative method of measuring the required exit criteria from the developmental reading course. Additionally, a beneficial side effect of the CPTs was felt by many of the staff of the College Preparatory Program. Their curiosity regarding the varied uses of computers was piqued, and when the hardware was not being used for testing, they explored other uses. The computer was integrated into the everyday course curriculum. Some staff members developed software that supplemented existing course curriculum; others purchased commercial software that augmented course materials. Some traveled to workshops and seminars to learn more about applications. The project and the explorations of the varied computer uses were positive experiences for students and staff alike.

Future Development

Given the positive experiences with computer adaptive testing, it is evident that such an innovation has potential for additional applications. In the near future, it is expected that Santa Fe Community College students will use computer adaptive testing throughout their college programs.

Largely as a result of the documentation presented to the state from the pilot test described here, Santa Fe has received preliminary approval to use the CPTs for state-mandated entry-level assessment and placement.

Further research is planned for next semester to investigate the placement of students within various levels of developmental courses. This placement, coordinated with curriculum revisions, will allow underprepared students to

remediate specific skill deficiencies in an expedient manner and advance to college-level work without the constraints of a traditional semester timeframe.

At the end of their college programs, all degree-seeking students are required to pass the state-mandated College-Level Academic Skills Test (CLAST) prior to earning an Associate of Arts degree. Initial efforts have been made by the state assessment office to investigate the use of computer adaptive testing for administration of the CLAST.

The combined development of computer adaptive testing and technology has the potential to become one of the more revolutionary accomplishments in education in this decade.

G. Thomas Delaino is Director, Management Information Services; Robert K. Myers is Executive Vice President; and Pat Smittle is Chair, Learning Labs and College Prep Program at Santa Fe Community College, Gainesville, Florida.

Reprinted with permission from *Computer Adaptive Testing: The State of the Art in Assessment at Three Community Colleges*, July 1988.

The Computerized Adaptive Testing Program at Miami-Dade Community College - South Campus

by Richard B. Schinoff and Larry Steed

Program Context

The College

Miami-Dade Community College is one of twenty-eight publicly supported community colleges in the state of Florida. It enrolls approximately one-fourth of all community college students within the state at its four campuses: North, Medical, New World Center, and South. The entire district enrolls approximately 45,000 credit students and 20,000 noncredit students during a fall semester. The South Campus, which uses adaptive testing for all students who are tested under the basic skills testing program, currently has 20,000 credit students and 10,000 noncredit students or just under one-half of the enrollment in the District. Hispanic students comprise about fifty-two percent of the enrollment, black students about five percent and white students approximately forty-three percent.

Over the past several years, Miami-Dade Community College has developed a student information system which is computer managed and readily accessible by students to assist in the admission, advisement, and registration process. The student information system at Miami-Dade includes an on-line application, registration and drop/add process. A graduation requirements document, the Advisement Graduation and Information System (AGIS) is also available on-line for student use. This particular student support program games out the courses completed on the student transcript, current registration and future registrations against the curriculum in effect for a particular student's graduation requirements. It also indicates courses that are required or suggested for transfer to one of thirteen universities in the State of Florida. All the support systems were in place at Miami-Dade that would allow new applicants to be processed immediately except for the testing program. The testing program was still a manual large group testing program with test booklets and bubble-in answer sheets.

The Missing Piece

Before the inception of the Computerized Adaptive Testing Program, a student would complete a Miami-Dade application and either mail it to the college or bring it to the Admissions Office; the student would then be sent away and told to wait for the next testing session. For example, the fall semester which begins in August had students applying to the college as early as the previous January, eight months before the term began. These students could not go through testing until June, six months after they applied for admission. Testing sessions were established for groups of one hundred individuals with three to five high-priced and specially trained testing proctors to administer the four paper and pencil subtests of the Florida MAPS examination for initial placement into English and mathematics courses.

Testing sessions did not occur year-round because of the high cost of proctors and also because each testing and orientation session took approximately three hours. A maximum of only three sessions could be offered during each day with a starting time of early in the morning, early in the afternoon, and very early evening. With the limited number of times that students could begin taking a test, many conflicts arose between work schedules and other personal commitments. It was believed that many students who applied to Miami-Dade never followed through with actual registration because the testing and orientation process was not convenient.

registration because the testing and orientation process was not convenient.

It was felt that if students could be tested for basic skills immediately after applying to the college, go through an orientation program, be advised and register, that the no-show rate would decrease. Even if the rate did not change, it was still believed that by offering the possibility of a student applying, registering, and paying fees within a three-hour timeframe any day of the week, year-round would certainly make the transition into Miami-Dade much easier.

Based upon the need to provide better services to students and the hope of reducing the no-show rate, the South Campus of Miami-Dade proceeded with the development of its Computerized Adaptive Testing Program. The IBM Corporation as part of its commitment to community college education agreed to provide personal computers for use in this pilot testing program. Likewise, the College Board agreed to provide its Computerized Placement Tests (CPTs) for use in the pilot program. As a result of this partnership between IBM, the College Board, and Miami-Dade, the Campus Assessment Center was established on the South Campus.

The Campus Assessment Center

Orientation and Assessment

Computer terminals were placed around the perimeter of the room with tablet armchairs in the middle of the room for use by other testing programs which were not administered in the computerized adaptive mode. Two orientation videotapes were developed for use in the assessment center. The first tape, approximately five minutes in length, provides information about the Computerized Adaptive Testing Program and the manner in which students use the computer terminal. This tape is shown to students before they take the adaptive test. The second tape, an orientation to South Campus' academic programs and student life, is viewed by the student after completion of the adaptive program. This tape is approximately thirty minutes in length and is viewed in conjunction with an academic planning manual. This allows the student to leave the orientation session with a manual in hand for academic planning purposes while at Miami-Dade. These tapes play continuously throughout the day with approximately fifteen earphones at desks in front of each monitor for student use. Therefore, there is no noise from the orientation program to interfere with testing.

After the student completes the test and before viewing the orientation video, the student calls the testing assistant to the work station to have the computer print the student's test results on the monitor screen. Currently, the testing assistant writes this information down on a piece of paper and enters the data into a terminal which is connected to the mainframe of the computer. The student is then asked to view the orientation video and upon completion of the video to proceed to the front of the testing room and pick up a Course Sequencing Pathways report.

Advisement and Registration

The results of the assessment, when combined together with the student's choice of transfer university and their program at Miami-Dade, generates the Course Sequencing Pathways document. This document is generated from a laser printer in the assessment center. The report combines data about the student's choice for transfer to one of the nine publicly supported universities in the state of Florida or to one of the four private universities in Dade County with assessment results to indicate the beginning level for English placement, math placement, as well as electives and general education courses that the student should take to satisfy the transfer requirements to one of the selected thirteen universities. Options also exist for the student who is undecided about a major or is undecided about transferring to a particular university. For the occupational/vocational student, who comprise about twenty percent of the South Campus population, a report will soon be available again indicating entry-level into English and math and outlining all the other courses the student is required to take to complete the appropriate Associate of Science or Planned Certificate Program of Study.

When the student leaves the Assessment Center with the course sequencing document in hand and having already completed the orientation program, the student is ready to meet with an advisor to discuss a program of study for the upcoming semester. Personnel in the assessment center direct the student to a specific department for advisement if the student has a declared major. If the student has an undeclared major, the student is referred to the Advisement and Counseling Department. After the completion of the advisement interview, the student is then able to immediately register for courses through the Touchtone telephone registration system at Miami-Dade. STAR Service (Student Telephone Assisted Registration) allows a student to register and pay fees by Master Card or Visa by using the key pattern on any touchtone telephone. The student may also stop by a Campus Bursar's Office to pay fees or mail a check to the Bursar's Office.

Continuous Student Intake

The Campus Assessment Center forms an integral part of the student flow model at Miami-Dade. It opens at 8:30 A.M. and allows the last student to enter the center at 6:30 P.M. The center then remains open until all students complete testing in the evening. Since the test and orientation program takes approximately two hours, the last student will most likely leave the Center before 8:30 in the evening. The Campus Assessment Center when used in conjunction with the online application process, the continuous orientation video program, the advisement interview, touchtone registration, and fee payments provides for a continuous on-line process of student intake. Since this process is continuous throughout the year, it is possible to better manage the enrollment process at the campus. Traditional peaks of heavy advisement at the end of the semester can be eliminated if student advising and registration are spread out during the entire term so that staff time can be better utilized. Advisors can spend more time with students if they come in earlier and do not bunch up at the end of the semester.

This is a massive re-education process which needs to take place so that students know they should register very early during a semester. Since early registration is encouraged, drops and adds are processed throughout the longer registration period as students' academic and personal needs change. For the fall semester, students who register before July 1, even as early as January, do not have to pay their fees until July 14. This is a real advantage for recruitment and enrollment management since Miami-Dade offers a complete orientation and advisement program, and gives the students their exact schedule with times and rooms that they will have for the fall term, many months before other universities and community colleges competing for the same student have even thought about their orientation and advisement programs.

Results of Pilot Administration

Publicly supported community colleges and universities in the state of Florida are required by law to administer or accept scores from one of four approved tests. Since the Computerized Placement Tests (CPTs) of the College Board were not on the approve list, Miami-Dade Community College, South Campus, received special permission from the Florida Commissioner of Education to administer the tests in lieu of one of the four approved tests.

During the summer of 1986, the CPTs were administered to students in three randomly selected classes, to some of the dual enrollees (students enrolling in college-level courses prior to completion of high school), and to a larger group of volunteers. These students had previously sat for the Florida MAPS test battery, and scores were available for both tests. The data collected (N = 265) were used as a supplement to data supplied by The College Board to establish a scale for the CPTs which approximated the scale scores for the four subtests of the Florida MAPS. This permitted use of existing computer files and the online Course Sequencing Pathways program with only slight modifications. The CPTs scores were coded as "Form P" for identification purposes, and were recorded on the existing mainframe file structure originally prepared for Florida MAPS. The scale was created with the intent of requiring the same proportion of students to enroll in college preparatory courses that the Florida MAPS series had generated in the past. It was an-

ticipated that slight modifications might have to be made when more complete data became available.

Reading Comprehension

Table 5.1 presents a comparison of the performance of all student tested on the Reading Comprehension Test from the MAPS series between March 1 and September 30, 1985 and 1986 with student performance on the CPTs during the same seven-month period in 1987. Although the populations tested in different years were similar, they were certainly not identical. However, performance on the MAPS exam has not varied significantly over the past few years. A comparison of the proportion of students failing the MAPS Reading Test in 1985 (34.5 percent) with 1986 (34.2 percent) demonstrates the relative stability of student performance in consecutive years. The passing score (scale score 12) is that established by the Florida State Department of Education. Degree-seeking students who score below 12 are required by public policy to enroll in college preparatory coursework within their first twelve credits. Students scoring in the range 12-14 are judged to be marginally passing. Student scoring in this range who enroll at Miami-Dade are advised to take a college-level reading course (REA 1105) to improve their reading speed and comprehension.

Based on the performance of the small group of students who had taken both the MAPS and CPT reading test in the summer of 1986, a raw score of 60 on the CPTs was originally selected as the score which would place the same proportion of students in college preparatory course work as a MAPS scale score of 12. When the CPT was administered to all students admitted for summer and fall terms, 1987, only twenty-four percent scored less than 60 while thirty-four percent of a similar population tested in 1985 and 1986 qualified for college preparatory reading on the basis of MAPS performance. It appears a CPT reading score higher than 60 is needed to identify students with deficiencies. Perhaps a score in the 64-65 range would have been a better choice.

Comparative Data

The Computerized Placement Tests were administered only on the South Campus at Miami-Dade. The remainder of the college continued to administer the Florida MAPS test battery throughout 1987. Tables 5.2, 5.4, 5.6 and 5.8 present pass/fail data for the MAPS test over the three-year period. Student performance appears to have improved slightly on three of the four MAPS subtests in 1987 over 1986. The percentage of students failing the MAPS reading test dropped from fifty-six percent in 1986 to fifty-three percent in 1987 as shown in Table 5.2. Whether this is an insignificant fluctuation or an indication that the 1987 south Florida high school seniors have greater academic skills than previous graduating classes is not known.

Writing

Table 5.3 provides pass/fail data for the objective writing tests on South Campus. The results for the Test of Standard Written English (the MAPS writing subtest) administered in 1985-86 are presented in the first two rows. The third row shows the performance of the 1987 population on the corresponding CPT. From 1985 to 1986 there was a slight decrease in the population of students failing the TSWE. This apparent increase in student performance in grammar, punctuation, and word choice was evident throughout the college. Table 5.4 shows this trend continued to a third year on those campuses administered MAPS through 1987. Based on this interpretation of the data, the passing score of 73, set for the trial period, identifies approximately the same proportion of students with writing deficiencies as a MAPS scale score of 30.

Arithmetic Skills

Due to a large number of students enrolling with serious academic deficiencies, Miami-Dade administers an arithmetic test to all entering degree-seeking students. Students with MAPS scale scores 101-115 are eligible for enrollment in an arithmetic skills course. The proportion of students scoring in this range has remained constant

over the past three years. A scale score of 116 identified approximately thirty-one percent of the students on South Campus as having deficiencies in arithmetic skills (Table 5.5, 1985-86) and approximately forty-four percent on the other campuses (Table 5.6, 1985-87).

During the trial period, a raw score of 57 on the CPTs arithmetic subtest was used as a minimum passing score for arithmetic. This score selected forty-one percent of the students for remediation rather than the thirty-one percent identified with a MAPS scale score of 115 or less, a difference of ten percent. Students scoring in this lower range in arithmetic but passing the algebra exam are not required to take an arithmetic course.

Algebra

Tables 5.7 and 5.8 demonstrate student performance on the MAPS exam as relatively stable in 1985-86 and improved slightly in 1987. In 1985-86, a MAPS scale score of 306 required elementary algebra of about forty-four percent of the South Campus students. The CPT raw score of 48 used as a passing mark in 1987 required elementary algebra of approximately fifty-five percent of the students. A CPT raw score in the vicinity of 41-43 would probably represent a more accurate estimate of MAPS Elementary Algebra performance at the 206 level.

Summary Results

All four subtests of the Computerized Placement Tests appear to be more discriminating between student performances on the lower end of the scale than the corresponding subtest on either of the former tests (MAPS and CGP) used for placement on South Campus in the past ten years. This observation was made when establishing a link between student performance on the CPT and the existing MAPS scale. If this is true, it may be possible to accurately place students in levels of college preparatory courses using CPTs test scores. This was not possible with the CGP and has not been possible with the Florida MAPS. The possibility will be examined further if access can be gained to student performance by items on the CPTs. The issue of placement validity will be also addressed in a future report.

Student Evaluation of the Program

Five days were randomly selected between March 1 and June 30, 1987, and students tested on these days were asked to evaluate the testing conditions. The form used consisted of fifteen items, each of which had from two to six possible responses to choose from. The students were asked to mark the most appropriate response to each of the fifteen items. Items nine to fifteen were printed on the back side of the one-page evaluation form. Thirty-two of the 369 students did not fill out the back side. The evaluation form was originally used at Santa Fe Community College in the summer of 1986 to determine the student opinions of computerized testing at that institution. The same form was used at Miami-Dade, with permission of the originator, so that comparisons could be made between student opinions from the two institutions located in different areas of the state with different demographic characteristics. Table 5.9 displays the survey questions and a summary of student responses.

Miami-Dade students liked the computer-administered test. Sixty-eight percent of the students preferred the computer test over the standard paper and pencil test, while only seventeen percent favored the traditional test. Fifteen percent of the students had no preferences indicated by responses to Item 13.

It is interesting to note that student opinions about difficulty of the verbal and mathematics portions of the exam (Items 1 and 2) were consistent with their performance on the test previously documented in Tables 5.1-5.4. Sixty-eight percent of the students felt the difficulty of the verbal tests was about right, while only forty-five percent felt the difficulty of the math tests was about right. Twenty-two percent felt verbal was too difficult while thirty-three percent felt math was too difficult. Very small numbers of students expressed an opinion that either part was too difficult or too easy.

Ninety-four percent of the students felt somewhat or very much challenged to do well (Item 3).

Forty-six percent felt the computer test was more challenging than paper and pencil, while only twelve percent felt the reverse (Item 4). About sixty percent of the students felt nervous about the test (Item 5), while forty percent felt that nervousness may have hindered performance (Item 6). A cross tabulation between these two items may be interesting. Student responses to Item 7 were a surprise. Only twenty percent of the students said they expected a computer test to make them more nervous, while forty-two percent expressed the opinion that a paper and pencil test would make them more nervous.

Sixty-one percent felt no time pressure (Item 8), and ninety-four percent of the students felt the computer was easy to use or very easy to use (Item 9). A large majority of the students felt the directions were clear (eighty-four percent), had no difficulty reading the screen (eighty-one percent), had a preference for a computer-administered test (sixty-eight percent), and felt it would be less tiring (fifty-nine percent). Ninety-five percent of the students felt the schedule hours were convenient, and finally, it was pleasant to notice that 100 percent of the students responding to the questions on page two judged the Campus Assessment Center staff as helpful.

From among the 3,000 people tested so far, only two have expressed a complaint about the computer-administered test. One was a professor reviewing the exam, and the other was a student. Both wore glasses with bifocal lenses and expressed concern that they could not position the screen at the right level to read it. A paper and pencil test was administered to the student, and the professor was informed that this option is always available to students, who, for any reason, express a need for a paper and pencil test.

Cost Comparisons

A comparison of the cost of the Computerized Placement Test with the cost of a traditional test must address three categories of expenses: personnel cost, compensation of proctors; royalties and fees paid testing publishers or distributors; and hardware costs, purchase, installation, and maintenance of computers, terminals and printers.

Personnel Cost

The personnel cost associated with the administration of the Computerized Placement Test from April 1 to August 31, 1987 is listed on Table 5.10. Table 5.11 presents the proctor cost of conducting the 93 individual sessions of MAPS over the same period of time in 1986. Part-time hourly personnel were used to proctor both types of examinations. The tables do not include the administrative costs associated with scheduling testing sessions, training and supervision of proctors, and preparing payroll rosters. These costs are comparable for the two programs. The costs presented in Table 5.11 include time allocated for proctor supervision of the video orientation. The cost of administering MAPS in 1986 does not reflect advisor time to conduct the group advisement sessions (approximately forty-five minutes per session). This time, formerly contributed by full-time professional staff, is now available for individual student counseling.

In comparing the $1.47 per student personnel cost for MAPS against the $1.88 per student for the CPT, it is important to note the differing philosophies behind the staffing and planning of the two different programs. The individual testing sessions scheduled in 1986 for the administration of MAPS were designed to minimize proctor costs. The primary directive for establishing the CPT schedule in 1987 was the convenience of the prospective student. During peak registration periods, the Campus Assessment Center was open five days per week and four evenings. This open schedule was established so testing could be scheduled and students could be instructed to return at a later time or new students applying to the college could take the test, go to advisement, then register for classes, all in the same day. In the past, students applying to the college were given a date to sit for the exam. Under the present policy, students newly admitted to the college are advised to report immediately to the Campus Assessment Center for the placement tests. A comparison of the costs between the two programs in the month of July ($1.45 vs. $1.49) illustrates, that with a large continuous flow of students, the cost differences between the two programs can be reduced to a minimal. It is also important to note that the rates

of pay for proctors in the two programs are not the same. Proctors who are required to give instructions to a large group of students (classroom situation) are paid at a rate of $8.55 per hour, while proctors who direct students on a one-to-one basis (laboratory type environment) are remunerated at approximately $5.25 per hour.

Test Royalties

Publicly supported colleges and universities in Florida are required by state law to administer one of the following entry-level tests to all degree-seeking students: ACT, SAT, ASSET, or Florida MAPS. Table 5.12 shows a comparison of individual test fees and the annual cost to administer each test to 8,000 students.

Students are usually required to register for national administrations of the ACT and SAT and pay the required fees several weeks ahead of time. Therefore, there is no direct fee cost to the local college. Information regarding tests administered on a residual basis was obtained by telephone from institutions using the tests and may not be exact. Residual testing offers greater flexibility to meet test requirements prior to registration, but with some loss of test security. These tests are also somewhat less expensive to administer.

The annual fees associated with testing 8,000 students range from a low of $2,000 for the Florida MAPS to a high of $96,000 for national administrations of the SAT. The SAT cost is based on a $12 per student registration fee. There is not a direct fee for the MAPS, however there is an ongoing cost of approximately $.25 per student associated with replenishment of test booklets and purchase of scanable answer sheets. In comparison, the projected cost of CPTs is $.75 per subtest, or $3.00 per student for the entire test battery. Use of a computer adaptive test such as the CPTs offers even more flexibility in scheduling than residual testing without the loss of test security.

Hardware Cost

The initial start-up cost for hardware required for the Computerized Placement Tests is substantial. To test 8,000 students per year on a walk-in basis (without appointments) would require availability of a minimum of twenty-five stations.

The cost per station varies considerably, depending upon the type of furniture and computer equipment utilized. A cost of $2,000 per station would represent a total capital outlay of $50,000. If the testing stations could be used for several testing requirements with peak use at different times, then the initial cost and a portion of the ongoing costs could be spread over the several applications to reduce per student cost.

Adoption of the Computerized Placement Tests as the primary entry-level test might appear to be very costly. It is interesting to note, however, that 8,000 students can be tested the first year at a per student cost ($11.13 including capital outlay for hardware) that is less than the registration fee for the SAT on a national test date.

Table 5.13 summarizes the first-year costs of using Computerized Placement Test for assessment at Miami-Dade Community College, South Campus.

Future Developments

The individual computers will be connected with a local area network so that the results of the Computerized Adaptive Testing Program can be transmitted to the mainframe automatically. This will eliminate the need for the testing assistant to extract manually the scores from the computer terminal, write them down on a piece of paper, and then enter them into the Miami-Dade computer system via a keyboard and terminal. With the implementation of direct transmission of scores to the mainframe, the testing personnel needed in the center can be reduced for a cost savings. Interactive video or laser disk orientation programs need to be developed that not only give information to students but also query the students to be certain they have learned key information from the orientation program.

Since there are many times during the semester when the dedicated computer equipment in the testing center receives low usage, the college purchased a software package which will allow Testing Services to develop and administer computerized adaptive tests for any faculty member who uses objective testing procedures. This last phase of development will relieve the faculty members from administering objective tests during precious class time which could better be used for classroom interaction to help students better understand materials, or to thoroughly review the results of the test.

Table 5.1
Reading – Frequency and Percent of Performance Classification on MAPS and CPTs Reading for Similar Populations, South Campus Data Only

	Performance Level			
Test and Dates of Testing	Fail	Marginal Pass	Pass	Totals
MAPS	1-11	12-14	15-25	
March 1-Sept. 30, 1985	1,955 34.6%	844 14.9%	2,851 50.5%	5,650 100%
March 1-Sept. 30, 1986	2,063 34.2%	909 15.0%	3,067 50.8%	6,039 100%
CPTs	20-59	60-68	69-120	
March 1-Sept. 30, 1987	1,505 24.1%	1,009 16.2%	3,727 59.7%	6,241 100%

Table 5.2
MAPS Reading – Frequency and Percent of Performance Classifications on MAPS Reading Tests for Similar Populations, College-Wide Data Without South Campus

	Performance Level			
Test and Dates of Testing	Fail	Marginal Pass	Pass	Totals
MAPS	1-11	12-14	15-25	
March 1-Sept. 30, 1985	3,085 53.7%	760 13.2%	1,903 33.1%	5,748 100%
March 1-Sept. 30, 1986	3,464 55.9%	848 13.7%	1,880 30.4%	6,192 100%
March 1-Sept. 30, 1987	3,110 53.1%	874 14.9%	1,874 32.0%	5,858 100%

Table 5.3
Objective Writing – Frequency and Percent of Performance Classification on MAPS and CPTs – Sentence Skills for Similar Population, South Campus Data Only

	Performance Level			
Test and Dates of Testing	Fail	Marginal Pass	Pass	Totals
TSWE	20-29	30-36	37-60	
March 1-Sept. 30, 1985	1,926	1,233	2,492	5,651
	34.1%	21.8%	44.1%	100%
March 1-Sept. 30, 1986	1,934	1,392	2,713	6,039
	32.0%	23.1%	44.9%	100%
CPTs	20-72	73-92	93-120	
March 1-Sept. 30, 1987	1,805	2,024	2,396	6,225
	29.0%	32.6%	38.4%	100%

Table 5.4
MAPS Writing – Frequency and Percent of Performance Classification on the Test of Standard Written English for Similar Populations, College-Wide Data without South Campus

	Performance Level			
Test and Dates of Testing	Fail	Pass	Marginal Pass	Totals
MAPS	20-29	30-36	37-60	
March 1-Sept. 30, 1985	3,091	1,120	1,536	5,747
	53.8%	19.5%	26.7%	100%
March 1-Sept. 30, 1986	3,209	1,300	1,674	6,183
	51.9%	21.0%	27.1%	100%
March 1-Sept. 30, 1987	2,886	1,303	1,660	5,849
	49.3%	22.3%	28.4%	100%

Table 5.5
Arithmetic -- Frequency and Percent of Performance Classification on MAPS and CPTs -- Arithmetic for Similar Populations, South Campus Data Only

Test and Dates of Testing	Performance Level: Fail	Pass	Totals
MAPS	101-115	116-125	
March 1-Sept. 30, 1985	1,763	3,885	5,648
	31.2%	68.8%	100%
March 1-Sept. 30, 1986	1,851	4,161	6,012
	30.8%	69.2%	100%
CPTs	20-56	57-120	
March 1-Sept. 30, 1987	2,574	3,759	6,333
	40.6%	59.4%	100%

Table 5.6
Arithmetic -- Frequency and Percent of Performance Classification on MAPS Arithmetic Tests for Similar Populations, College-Wide Data Wwthout South Campus

Test and Dates of Testing	Performance Level: Fail	Pass	Totals
MAPS	101-115	116-125	
March 1-Sept. 30, 1985	2,513	3,293	5,806
	43.3%	56.7%	100%
March 1-Sept. 30, 1986	2,721	3,405	6,126
	44.4%	55.6%	100%
March 1-Sept. 30, 1987	2,546	3,274	5,820

Table 5.7
Algebra – Frequency and Percent of Performance Classification on MAPS and CPTs – Algebra for Similar Populations, South Campus Data Only

Test and Dates of Testing	Performance Level: Fail	Marginal Pass	Pass	Totals
MAPS	201-205	206-209	210-225	
March 1-Sept. 30, 1985	2,541 43.4%	893 15.2%	2,424 41.4%	5,858 100%
March 1-Sept. 30, 1986	2,736 43.8%	1,004 16.1%	2,507 40.1%	6,247 100%
CPTs	25-47	48-63	64-120	
March 1-Sept. 30, 1987	3,594 55.5%	935 14.4%	1,951 30.1%	6,480 100%

Table 5.8
Algebra – Frequency and Percent of Performance Classification on MAPS Algebra Tests for Similar Populations, College-Wide Data without South Campus

Test and Dates of Testing	Performance Level: Fail	Marginal Pass	Pass	Totals
MAPS	201-205	206-209	210-225	
March 1-Sept. 30, 1985	3,564 61.4%	846 14.6%	1,396 24.0%	5,806 100%
March 1-Sept. 30, 1986	3,750 60.0%	811 13.0%	1,687 27.0%	6,248 100%
March 1-Sept. 30, 1987	3,334 56.8%	813 13.8%	1,723 29.4%	5,870 100%

Table 5.9
Percent of Student Responses – South Campus Assessment Center Student Evaluation, March-June, 1987

1. How difficult was the verbal part of the test for you? (that is, Reading and/or Sentence Skills)

2%	much too difficult
4%	somewhat too easy
22%	somewhat too difficult
1%	much too easy
68%	just about right
3%	did not take either of the verbal parts

2. How difficult was the mathematics part of the test for you?(that is, Arithmetic and/or Algebra)

9%	much too difficult
10%	somewhat too easy
33%	somewhat too difficult
3%	much too easy
68%	just about right
–	did not take either of the verbal parts

3. Did you feel challenged to do as well as you could on the test?

7%	not at all
55%	somewhat
38%	very much

4. In comparison with a paper and pencil test, did you find the computerized test...

46%	more challenging than a paper and pencil test
42%	about the same as a paper and pencil test
12%	less challenging than a paper and pencil test

5. Were you nervous while taking the test?

40%	not at all
48%	somewhat
12%	very much

6. Did nervousness while taking the test prevent you from doing your best?

9%	yes, definitely
31%	yes, somewhat
37%	probably not
23%	definitely not

7. Which would make you more nervous, a paper and pencil test or a computer administered test?

42%	a paper and pencil test
20%	a computer administered test
38%	they'd be about the same

8. Did you feel any pressure to complete the test quickly?

6%	yes, a great deal of pressure
33%	yes, some pressure
61%	no

9. How easy was it to use the computer to take the test?

75%	very easy
20%	reasonably easy
5%	somewhat difficult
–	very difficult

10. Were the directions given by the computer clear and easy to follow?

84%	yes, very clear
15%	reasonably, clear
1%	no, not very clear

11. Did you have any difficulty reading the material presented on the computer screen?

9%	yes
81%	no

12. In comparison with a paper and pencil test of the same length, did you find the computer test...

10%	more tiring than a paper and pencil test
31%	about the same
59%	less tiring than a paper and pencil test

13. Which would you prefer to take?

68%	a computer administered test
17%	a paper and pencil test
15%	do not prefer one over the other

14. Were the hours that the computer testing lab was open convenient times for your participation?

95%	yes
5%	no

15. Was the computer testing lab staff helpful?

100%	yes
< .5%	no

Table 5.10
Proctor Cost for Administration of Florida MAPS, April-August, 1986

Month	Number of Sessions	Proctor Hours	Cost @ $8.55/hr.	Number Students Tested	Cost per Student
April	14	147	$1,257	800	$1.57
May	6	63	539	350	1.54
June	12	126	1,077	873	1.23
July	31	326	2,783	1,915	1.45
August	30	315	2,693	1,754	1.54
Totals	93	977	$8,349	5,692	$1.47

Table 5.11
Proctor Cost for Administration of Computer Placement Test, April-August, 1987

Month	Proctor Hours	Cost @ $5.25/hr.	Number Students Tested	Cost per Student
April	418	$2,194	891	$2.46
May	375	1,969	650	3.02
June	416	2,184	1,237	1.77
July	470	2,467	1,661	1.49
August	556	2,919	1,797	1.62
Totals	2,235	$11,733	6,236	$1.88

Table 5.12
Cost of Entry Level Tests in Florida

Testing Program		Administration Fee or Materials (per student)
ACT		
	National Administration	$10.50
	Residual Basis	8.00
ASSET		2.50
CPT		3.00
MAPS		.25
SAT National Administration		2.00

Table 5.13
First Year Cost of CPTs at Miami-Dade - South Campus

	Per Student	Total Cost 8,000 Students
Personnel Cost	$1.88	$15,000
Test Usage Fee	$3.00	$24,000
Hardware	$6.25	$50,000
Total	$11.13	$89,000

Reprinted with permission from *Computerized Adaptive Testing: The State of the Art in Assessment at Three Community Colleges.*

Richard B. Schinoff is Dean of Students, South Campus; and Larry Steed is Director of Testing Services, South Campus at Miami-Dade Community College, Miami, Florida.

A Summary of Insights From Experience:

Guidelines for Implementing Computerized Adaptive Testing for Assessment

The previous chapters have described the experiences of three community colleges in implementing computerized adaptive testing for entry-level assessment. These chapters provide numerous insights to other colleges contemplating the use of The College Board's Computerized Placement Tests or other tools that might be available, and to a lesser extent, these experiences suggest how a community college might go about using the technology of computerized adaptive testing for purposes other than assessment for entry-level skills courses.

All three of these community colleges had implemented assessment and placement policies and procedures prior to experimenting with computerized adaptive testing. As a result, their experiences will be most pertinent to colleges seeking to convert from paper and pencil assessment testing to computerized instruments. Colleges considering implementation of computerized adaptive testing as the technological basis for their first assessment system should also find these colleges' experiences to be relevant to their purposes. However, the perspective of these colleges is that computerized adaptive testing using Computerized Placement Tests represents the state of the art in assessment, with benefits far exceeding costs. Therefore, the insights gained from their experiences do not include consideration of the relative merits of other types of assessment instruments that a college first implementing an assessment program might consider.

So, the following is a summary of the insights gained from the experience of Central Piedmont Community College, Miami-Dade Community College, and Santa Fe Community College implementing computerized adaptive testing for assessment. To whatever extent they are relevant, they represent guidelines for other colleges seeking to implement the same technology.

Planning and Involvement

1. Implementing computerized adaptive testing for assessment requires the same effort to gain acceptance among important college constituencies as implementing any change or innovation. The whole-hearted commitment of the college's leadership must be secured. Affected staff must be involved in planning and implementation, and various college constituencies must be persuaded of the benefits to them, to students, and to the college of implementing the new system.

2. It is the responsibility of those leading the implementation of computerized adaptive testing to educate the leadership of the college, faculty, counselors, and other staff about the innovation. Professionals in the testing field are helpful to explain how computerized adaptive testing works, how it relates to developments and trends in testing, and its benefits. Demonstrating how computerized adaptive testing works in sessions that allow affected faculty and staff to experience testing procedures, ask questions, and express concerns can be effective in reducing anxiety and resistance inherent in implementing any change.

3. The relative benefits of replacing a paper and pencil-based assessment system with the CPTs, depend somewhat on the characteristics of the

existing system. A persuasive case can be made that the principal benefits of using computerized adaptive testing for assessment include the following:

a. CPTs increase flexibility by allowing both on-demand testing and scheduled testing for both individuals and groups.

b. CPTs increase accuracy in assessing students' entry-level skills for the entire range of skills covered by the instrument (whereas some tests discriminate poorly either at the upper or lower ends of student skills).

c. CPTs provide immediate scoring and feedback to students, advisors, and counselors.

d. CPTs reduce student frustration with testing by generating appropriate questions based upon a student's previous responses.

e. CPTs are untimed and reduce test anxiety.

f. CPTs can be completed in roughly one-half of the time of comparable paper and pencil tests.

g. CPTs scores can be equated to other testing instruments and comparable cut-scores for placement in appropriate courses can be established.

h. Like other standardized tests, normative data for CPTs are available for comparisons and other uses.

i. As a result of its features, CPTs may assist recruitment and reduced attrition during the enrollment process, and may increase student success in appropriate courses.

j. Initial costs in hardware may be offset by increased efficiencies, including reduced costs for test administration. Also, hardware can be made available for other testing or instructional uses.

4. Existing assessment and placement policies and procedures must be evaluated to determine changes needed to implement the new technology and to make effective use of its unique features. The effect of these changes on staff and on other college procedures and policies must also be carefully evaluated.

5. It is advisable to run computerized adaptive testing initially as a parallel system to the existing assessment process to allow a smooth transition to then new system, to minimize the consequences of error in initial implementation, and to minimize staff resistance to changing to the new system. The benefits of computerized adaptive testing will be apparent and persuasive in comparison to existing paper and pencil procedures.

Designing an Appropriate Facility

6. It is important that the space designated for computerized adaptive testing be easily accessible to all students. The space should be in reasonable proximity to other student intake functions, such as advisement, registration, and counseling. It is also desirable to locate the space used for computerized adaptive testing near existing testing facilities to increase efficiency and to take advantage of student familiarity with existing services.

7. The facility must provide adequate security for costly hardware and software.

8. The facility must be large enough to accommodate the equipment and furnishings while allowing students to enter and exit without disturbing other students. It must be arranged to allow students and testing personnel to move about easily and to provide sufficient space to accommodate handicapped students in wheelchairs. Computer stations arranged around the perimeter of a room facing the walls appears to be an arrangement that works well to accomplish these objectives.

9. Although windows are attractive, bright light can create a nuisance by causing glare on the computer screens. If the facility has windows, good blinds or shades need to be installed to control lighting problems.

10. The facility must be well-ventilated. Normal classroom ventilation systems often need to be modified by the addition of ducts and fans to

accommodate the heat generated by the computer equipment and a room full of students.

11. The facility must be adequately wired to provide safely and conveniently for the large number of electrical outlets required by the equipment. If the space is to accommodate a network of microcomputers, arrangements must be made for the required cable. If feasible, the facility can be "hard wired," that is, with customized power and cable hookups built into the walls and floor to accommodate the design of the testing facility.

12. Since the testing system is dependent upon electrical power, a back-up source of power to protect the systems from electrical surges or momentary losses of power (uninterrupted power suppliers) are recommended if feasible. At a minimum, procedures need to be developed in case of a power failure.

Equipping and Furnishing the Facility

13. The number of computer stations with which the facility is equipped may be prescribed by available physical space and financial resources. One college provided adequately for its students' testing needs using five microcomputers with scheduled appointments. A testing facility equipped with twenty to twenty-five stations can accommodate the needs of most large community colleges, while also providing the capability to conduct testing for scheduled groups or to use the facility for instructional purposes with class-size groups.

14. The number of stations required to equip a center for adaptive testing for assessment is determined, in large part, by the planned testing schedule. Testing by appointment requires the fewest stations. On-demand testing of individuals requires more stations than scheduled appointments, especially during peak registration periods. Group testing requires one station for each member of the group.

15. While equipping each computer with its own printer can be convenient for students, both the cost of purchasing multiple printers and the noise created by several contact printers in operation simultaneously recommend against such an arrangement. A single printer ported to a file server supporting a network of all test stations is probably the best arrangement. Providing one or more printers in an adjacent work area, or to be used in the testing facility during non-testing times, is an acceptable compromise.

16. Both tables and individual carrels can be used to support the computer stations. Tables provide a more open environment than carrels, take less space, are more flexible for other uses, and are less expensive. Since computerized adaptive testing generates original tests for each student, carrels are not required to reduce copying. Comfortable, easily adjustable chairs are a necessity.

Scheduling and Using the Facility

17. The scheduling of the facility is determined in large part by its capacity. With a limited number of computer stations available, scheduling appointments for individual students probably best utilizes the equipment – with drop-in students accommodated as possible. The ideal arrangement is to provide for both scheduled appointments and on-demand testing.

18. It is important to provide testing services not only during prime daytime hours, but also during evenings and on weekends when students are present. To take full advantage of the testing technology, testing services should be available during the same hours that other student services, such as advising and counseling, are available.

19. One of the advantages of computerized adaptive testing is to provide on-demand testing throughout the year, not just during registration periods. Services should be provided year-round to take advantage of this feature, though daily hours of operation might be adjusted to reflect peak registration periods or other scheduling demands.

20. Given the investment in the computer equipment required to implement computerized adaptive testing, other applications should be considered for the facility and equipment. The facility can be used effectively for a variety of testing applications other than entry-level assessment, including ad-

ministration of course tests. Careful scheduling can allow the facility to be used to provide instruction to students, as well as to provide staff development opportunities to faculty. However, it is important to assure that the scheduling of the facility for group testing or for other purposes does not compete with its principal purpose – to provide year-round, on-demand assessment testing.

Staffing and Orientation

21. One person should be designated to coordinate facility operations. The person must be a competent manager who is student-centered and has a strong customer-service orientation.

22. Sufficient numbers of support staff required to provide adequate coverage and back-up for operating hours need to be hired and trained. The staff must be trained in the use of microcomputer hardware and peripherals, in testing center operations and procedures, and in the use of the testing software. To assist in implementing the CPTs, staff must be familiar with the CPTs Coordinator's Notebook and must learn how to operate both the test and management software (CPMS).

23. Orientation sessions should be conducted for faculty, advisers, counselors, and other staff affected by implementing the technology. Providing staff with the opportunity to participate in hands-on simulation of the testing system "as students" not only provides staff with an appreciation for both the technology and the students' experience, but also allows staff to identify bugs in the system or procedures that can be corrected prior to full-scale implementation.

24. Testing center staff should conduct several trial runs of the system prior to implementation, including all aspects of the system from orientation sessions to dissemination of test results.

25. Efforts should be made to acquaint students with the testing procedures and technology in order to reduce their apprehension of the new testing situation. Counselors, advisers, faculty, and other staff must be sufficiently familiar with the system to be able to provide students with appropriate assurances. A five-minute videotape can be developed and shown to students to introduce them to computerized adaptive testing.

26. Students should be provided with as much assistance as possible to use the testing tools effectively. For example, one college found that placing brightly colored dot stickers on the three principal function keys used to administer the CPTs greatly increased the ease with which students learned to operate the testing equipment. Testing assistants should be available to help students operate the hardware and software at all times.

Reporting and Storing Test Results

27. The simplest way to disseminate test results is to provide a score report for the student immediately upon test completion. This can be taken to an appropriate advisor, counselor, or faculty member for course placement or further interpretation. This procedure can be facilitated with the use of multi-part forms generated by the computer.

28. Institutions with more sophisticated student information systems that provide counselors and advisors with on-line access to student records and other pertinent program information can immediately upload test results into the mainframe student information and advising system. Procedures for uploading results include manual data entry from a printed report using a conveniently located terminal, as well as developing customized programs that automatically update mainframe records with data supplied by the microcomputer-based testing software.

29. Institutions may also consider developing reports that are customized to combine test results with other student information and diagnostic data to provide precisely those data elements needed to operate existing college processes. It is critical to the successful implementation of computerized adaptive testing that it is incorporated into other student-intake processes. It is possible to customize test reports to assist such integration.

30. Test scores may also be entered into mainframe student files daily using batch processing of test

results from hard copy individual or cumulative reports provided by the software.

31. Regardless of the way in which test scores are reported and disseminated to affected offices, procedures must be developed to maintain electronic data files in either the mainframe or microcomputer that are accessible by student, date, and test. Hard copy reports of test results may be maintained for back-up documentation, but to reduce paper storage requirements, procedures should be developed to determine how to dispose of historical reports on a regular schedule.

32. It is important to build into any reporting and dissemination procedures for test results appropriate levels of security, for test results are confidential student information.

Using Test Results

33. The use of the results of assessment is a policy decision that has far-reaching consequences for a college's instructional and student services programs. These policy decisions are independent of the technology used to administer assessment testing for entry-level skills. To justify the investment in resources, however, the results of computerized adaptive testing probably need to be integrated into advisement and course placement processes.

34. In the case of a college replacing paper and pencil assessment testing with computerized adaptive testing technology, the use of test results for advisement and course placement will have already been established by policy. If cut-scores for placement in courses have been previously established and validated, comparable cut-scores for Computerized Placement Tests can be developed.

35. The College Board and ETS have equated scores from the CPTs with other common standardized tests. Similarly, normative data are available to provide comparisons with student scores at various types of institutions. These data provide a useful starting point for determining cut-scores that effectively place students in appropriate courses. Nonetheless, cut-scores should be determined and modified as necessary to meet the needs of their students and curriculum by each institution using the CPTs.

Evaluation

36. A substantial plan to evaluate computerized adaptive testing for assessment into entry-level skills courses must be developed prior to initial implementation. The principal purposes of the plan must be to determine the effectiveness of test scores in placing students in appropriate classes and to solicit student and faculty feedback on testing software and procedures.

37. A research plan to assess the effectiveness of CPTs scores in placing students in appropriate courses requires systematic collection of student and test data, including follow-up of student performance in subsequent classes. Results of an initial implementation of CPTs can be compared to results produced using previous assessment instruments. Similiarly, results of parallel assessment systems can be compared to evaluate the relative effectiveness of CPTs for placement into courses.

38. Testing staff should involve other college staff with research expertise—generally from the institutional research function—in the design and implementation of the evaluation.

39. The evaluation of the effectiveness of test scores in placing students in courses should be conducted semi-annually, and results should inform the decision to modify cut-scores as necessary to improve effectiveness as determined by institutional goals. Normative data can also inform cut-score determination. After the experience of determining and using cut-scores over several semesters, the need for modifications should be reduced.

40. The best source of formative data on testing instruments and procedures are student evaluations. A short evaluation form should be developed and administered to students upon completion of testing. Initially, all students might be asked to complete the form until the most obvious procedural problems are solved; after a semester or more of experience, the evaluation

can be administered to random samples of students.

These recommendations are the result of the initial experiences of three community colleges using computerized adaptive testing for assessment. All three used the most current version of The College Board's CPTs available at the time, and all three have continued to learn from their experiences and to modify their systems to improve their effectiveness in using this testing technology for assessment and other purposes. CPTs have also been improved since their initial introduction. The most recent version (2.0) also provides the capability of running an assessment testing system on a network of microcomputers. Also, the research on test results has continued to provide more complete normative data and better information to colleges contemplating using the CPTs. While these recommendations do not reflect the continued improvements being made in computerized adaptive testing for assessment purposes, they do provide helpful insights gained from pioneering work in the field.

Reprinted with permission from *Computerized Adaptive Testing: The State of the Art in Assessment at Three Community Colleges,* League for Innovation in the Community College.

Assessment's Next Wave: The Computerized Placement Tests

by Pat Smittle

For the purposes of efficient advisement and appropriate placement in course work, accurate assessment of basic skills is a major concern for most postsecondary institutions, especially those that have open admission policies that provide access for all students, without regard to their academic preparedness. At some point, this concern forces each of these institutions to examine the relationship between various tests it uses and the institution's specific mission and student population.

In 1986, Santa Fe Community College in Gainesville, Florida, in cooperation with the College Board, IBM, and Miami-Dade Community College, initiated a project to investigate the feasibility of using the Computerized Placement Tests (CPTs) for entry-level assessment and placement. The project was sponsored by the League for Innovation in the Community College. Developed by the College Board, the CPTs constitute a new testing program that is based on the theory of computerized adaptive testing. Computerized adaptive testing is a method of testing that utilizes the capabilities of microcomputers to adapt examinations, by level of difficulty, to the abilities of individual students. Instead of facing a battery of prearranged test questions, many of which are either too hard or too easy, the computer quickly presents students with questions that are appropriate to their ability and knowledge.

The CPTs consist of four untimed tests that cover reading comprehension. sentence skills, arithmetic, and elementary alebra. A fifth test – college-level mathematics – is undergoing field trials and will be introduced this September. The issue of entry-level testing is a critical one for those colleges involved in the project because the Florida Community College system adheres to an open-admission policy and promotes access and excellence for all students. To assure academic excellence for the underprepared students who are accepted, Florida legislation mandates that every first-time-in-college student who is seeking a degree shall take an assessment test, register for basic skills courses if deficiencies are indicated, and demonstrate designated competencies before enrolling in most college-level courses.

Since more than 50 percent of the incoming freshmen at Santa Fe Community College are enrolled in at least one developmental course, our initial exploration of the CPTs focused on this area. To collect data and gain experience with testing procedures that were quite different from the traditional paper-and-pencil tests, the CPTs were administered to those students who had been placed in developmental reading, writing, and mathematics courses primarily on the basis of their ACT or SAT scores.

Soon after the project began, it became obvious that students liked the "new" test. Student evaluations of the CPTs indicated an overwhelming preference for the CPTs over paper-and-pencil tests; 63 percent of the students reported that they preferred the computer test. Students reported that they thought the computer test was very easy, even those whose skill deficiencies were still manifest.

It was this initial, enthusiastic reaction that piqued our interest and encouraged us to continue our exploration. The initial assumption was that this positive reaction was due to the novelty of using the computer. However, after approximately three years of exploration with the CPTs and observation of students who use computers on a regular basis, it has become apparent that satisfaction with the test is based on factors other than novelty. Among students taking basic skills courses, we wanted to explore what it was about them that might explain their positive reaction to the CPTs, as contrasted with their usual fearful reaction to tests.

Underprepared Students and CPTs

In our work, we found that several characteristics of underprepared students make them likely candidates for administration of a CPT. For one thing, the short attention span of many developmental students creates an automatic barrier to the traditional paper-and-pencil tests that are usually designed to assess a wide range of abilities. According to William C. Ward, senior development leader at Educational Testing Service, this "traditional" type of test usually consists of some easy and difficult items, with most of the items being at a moderate level of difficulty. It is a rigid structure that requires students to respond to many items at all degrees of difficulty to determine their level of achievement. The computerized adaptive test, however, is individualized, having a fluid structure that depends on a student's response to each question. This allows the level of achievement to be determined in a short time, using only a few items in the general ability range of each student.

Since the CPTs are adjusted for each student and since that student works only with questions in his or her ability range, the element of discouragement is diminished, allowing underprepared students to demonstrate what they know instead of what they do not know. A statistical procedure called Item Response Theory, which is one of the critical components of computerized adaptive testing, estimates the individual's ability based on the difficulty of the questions the individual can answer correctly. The procedure allows us to determine achievement levels, with items presented at whatever level of difficulty is appropriate for the student.

Developmental students perform optimally when given simple and specific instructions. The introduction and instructions for the CPTs are easy to read and very clear. After students enter personal information such as name and social security number, they are instructed to use only three different keystrokes throughout the entire test. Instructions are so clear that little additional information has been required by our test administrators, regardless of a particular student's level of familiarity with the computer. Instructions further inform the students that the tests they are about to take are untimed, eliminating the pressure factor.

Although it is not always evident, many developmental students at the postsecondary level have low self-esteem, especially in relation to academics. But when students are seated before computers that are used to administer the CPTs, they infer that they are capable of working with the same high technology equipment that "successful" people use. And while feelings of distraction often characterize many developmental students, the computer demands a focused interaction from them that is generally forthcoming, perhaps because students know they have to focus only for a short period and they feel challenged by the presentation of questions at appropiate ability levels.

At the beginning of the test, students are informed that they will be given their scores upon completion. As soon as the tests are completed the computer calculates the scores and displays them on the screen. This immediate feedback is another important element in raising the self-esteem of developmental students.

The computer eliminates another problem particular to underprepared *adult* students: embarrassment. Though many students are self-conscious about their deficiences in basic skills, the computer's immediate scoring capability removes the "human" element that students often view as judgmental.

Developmental students thrive on success. Regardless of the level of achievement, every student who completes the CPTs has been successful with high technology equipment. This successful experience, although small, provides an incentive to move to the next step in one's college career. All students need reinforcement, but it is critical for underprepared students who have experienced repeated failure in their academic work. Reinforcement is the "green light" that tells the student to keep going. The test-retest capabilities of the CPTs allow for frequent testing to reinforce the efforts of the students themselves. These capabilities also make the tests valuable measures of progress. The minimal cost of 25 cents for each retest makes the CPTs an inexpensive motivator for students and a counseling and advisement tool for the instructor.

Considering the similarities between characteristics of developmental students and the attributes of the CPTs, it is not surprising that developmental students respond well to the tests. It should be noted that included in these common characteristics are the basic principles of individualized instruction that Pat Cross identifies as being essential ingredients for effective learning. These individualized tests have all the necessary components to become a vital part of the entire instructional program for underprepared students.

Conclusion

The large numbers of underprepared students admitted to open-admission institutions demand that we actively pursue nontraditional methods of meeting the needs of our at-risk population. Given our limited resources in terms of time and money, it behooves us to strive for the most efficient and effective assessment and instructional materials available. Our objective is to admit these students, regardless of their level of academic preparedness, and remediate their deficiencies, thus preparing them for college-level courses that aspire to high academic standards. Obviously, the first step for success is an accurate assessment of their skills, so they can begin their work at the appropriate level.

At Santa Fe Community College, we have recently made another advancement in meeting the needs of underprepared students who are admitted to our college. We have adopted the CPTs as our on-site assessment instrument. Although our students were major factors in our adoption of this particular test. it would be wrong to conclude that the CPTs are not appropriate for higher level students also. Generally, students with good skills perform well on most tests. (Our extended project included students at all educational levels, and we continue to see positive results from all of the students at all different levels.)

Our many years of experience with underprepared students demonstrate that those who are placed at the appropriate level of academic work upon their entrance to college can develop the skills necessary to earn a degree. Longitudinal studies reveal that approximately 30 percent of the graduates at Santa Fe Community College were placed in at least one developmental course at the beginning of their college career. This is significant when we acknowledge that, due to basic skill deficiencies, these students would not even have been admitted to some institutions that have selective admission policies.

In 1983, Howard Wainer, senior research scientist at Educational Testing Service, wrote. "We are on the verge of a technological revolution in testing as a result of the availability of extensive, inexpensive computing and some recent developments in statistical test theory." In 1990, I would like to expand on that thought: We are in the

midst of a technological revolution that is moving testing into a new era. The computerized adaptive mode provides, for test administrators and test takers, an efficiency and effectiveness previously unavailable. The computerized placement tests are, indeed, assessment's "next wave."

The CPTs at Santa Fe Community College

Before switching from the traditional paper-and-pencil assessment which Santa Fe Community College had used since 1981, a study to determine the effectiveness of placement was conducted. Five-hundred-thirty-nine students were randomly assigned to take the ACT or the CPTs. The resulting data provided the basis for establishing a relationship between the two tests. Also. analysis of the relationships between entering students' test scores for the ACT the CPTs and final course grades provided information relative to the predictive ability of each test.

In all courses except one, the correlations between the CPTs scores and the final course grades were greater than the correlations between the ACT scores and the final course grades. Only in the developmental writing course was the relationship with the ACT slightly higher.

Since the efficiency of the CPTs was already apparent, the information was used to answer the most critical question: "Which test is the best predictor of success in the courses in which the students are placed?'

It was determined that the CPTs are better predictors of success than traditional paper-and-pencil tests. Last winter, the CPTs were therefore adopted as the on-site basic skills assessment tools for Santa Fe Community College.

Computerized Adaptive Testing:
It's Efficient, It's Flexible, It's Fun

At last, a test that gets high marks from administrators, faculty, *and* students. The Computerized Placement Tests (CPTs) are popular among administrators for their efficiency and flexibility. They're popular with faculty, since the accurate course placement that derives from the test results supports instruction. And students like CPTs for the sensitive, individualized assessment they provide.

CPTs use computerized adaptive testing, a measurement and programming technology that makes branching decisions – decisions that allow the assessment to adapt to each student.

The CPT taker is first presented with a question of moderate difficulty. If it is answered correctly, a very difficult question appears; if answered incorrectly, a very easy question appears. As the test progresses, it pinpoints the student's skill level. The test score indicates the highest difficulty level reached.

CPTs are extremely efficient because students do not spend time with questions that are far too easy or too difficult for them. The test conforms to the test taker, rather than the other way around.

Because of their efficiency and the flexibility of computer administration. CPTs have no time limit. This is a popular feature among students who perceive that an assessment without time pressure is more accurate. The median time spent is approximately 20 minutes per test – 30 to 50 percent less time than a comparable paper-and-pencil test.

Because CPTs are adaptive, each form of the test is different: tests are highly secure and retest costs are low.

Colleges administer CPTs on campus in a variety of testing arrangements, using single, stand-alone computers, a laboratory of 20 or more stations, or terminals in several locations connected by a network.

Each CPTs draws from a pool of 120 questions. Each test comprises a fixed number of questions (from 12 to 20, depending on the subject).

CPTs are currently offered in Reading Comprehension, Sentence Sense, Arithmetic, and Elementary Algebra. College-Level Mathematics will be introduced in September.

When students sit down at a terminal to take a CPT, before actual testing begins, they are given a brief run-through to familiarize them with the few keyboard commands they'll have to use. They also enter identifying information and answer background questions.

Based on this information, colleges can flexibly configure their CPT system or network to call up appropriate tests. For example, students who indicate that they have had a background in college preparatory mathematics might be given the College-Level Mathematics module. They wouldn't have to take Arithmetic or Elementary Algebra as they would if they were using a test booklet.

Computerized scoring provides immediate reports for student and college use.

Using Computerized Placement Management System software, a college's placement decision rules, students' CPTs scores, and other information can be integrated. Incoming students can go to a one-stop assessment center to take the CPTs, receive an instant printout of their course placements, and on-site academic advising.

The Placement Research Service provides an annual validity analysis to assure accuracy and consistency of placement.

Program enhancements include Seamless Serial Testing, which allows questions to be selected from two or more pools of items that are unknown to the user. In addition, several high school-college collaboratives are being piloted.

CPTs are in use at a growing number of colleges across the United States and Canada, where they are proving to be a flexible and economical assessment that students adapt to as readily as the tests adapt to them.

References:

Cross, Pat. *Accent on Learning.* San Francisco: Jossey Bass, 1979.

College Board Computerized Placement Tests (CPTs): Computerized Adaptive Testing Progam for College Placement. Princeton, New Jersey: Educational Testing Service and the College Entrance Examination Board, 1986.

Computerized Adaptive Testing. Laguna Hills, California: League for Innovation in the Community College, 1988.

Wainer, Howard. "On Item Response Tehory and Computerized Adaptive Tests: The Coming Technological Revolution in Testing" in *The Journal of College Admissions*. Vol. 28, No. 4., April 1983.

Ward, William C. "Using Microcomputers to Administer Tests." In *Educational Measurement: Issues and Practices*. Summer 1984, 16-20.

Reprinted with permission from *The College Board Review*, No. 156, Summer 1990.

Statistical Analysis of the Class of 1994 at the USMMA

by Charles Weber

Scope and Purpose

Starting with the Class of '90 and continuing through the Class of '93, the U.S. Merchant Marine Academy gave the Elementary Algebra Computerized Placement Tests (CPTs) to the incoming students during Indoctrination (a two-week orientation period before classes begin). The results have been used to choose the students who will be placed into our so-called "extended" sections, taking an extra day per week of mathematics devoted to algebra and trigonometry remediation. This CPTs test improved the prediction of first quarter calculus grades, compared with the Mathematics Scholastic Aptitude Test (MSAT) alone, but it was inadequate for two reasons. First, the test is in elementary algebra and therefore does not cover all the mathematical preparation that is necessary to do well in calculus. Second, our students generally do quite well on the test, bunching the grades toward the upper end. There was insufficient discrimination between weak and strong students. In other words, many weak students (students at high risk of failing calculus) did fairly well on the CPTs Elementary Algebra. The reason for this is, of course, the limited coverage of the test.

We have, for the first time this year, given our entering class (Class of '94) a new CPTs test, called the College-Level Mathematics test (CLM). This report is an analysis of the results of that test from two points of view: First, how good a predictor of calculus grades (or even QPA [Quality Point Average]) is the new test? Second, what other information can we obtain from the results of the test?

Comparing the Preparation of Different Entering Classes

The major measure of the mathematical ability of incoming students that is available to the Admissions Office is the MSAT. We have accumulated ample evidence in recent years at the Academy (and have heard similar comments from people at other colleges) that the MSAT is inadequate as a predictor of success in calculus (although it might be a better predictor of overall success in college). (Since these observations are based only on students who have been accepted into college, the low correlations do not mean that the tests are not effective in screening out students with little chance of success.) Analysis of first quarter calculus grades for the Classes of '90 through '94 shows the correlation coefficient (r) to vary from a high of .53 with the Class of '90 to a low of .20 with the Class of '93. It has averaged .38 – rather low. Incidentally, the correlations of the Verbal SAT and High School Rank with Calculus I were .15 and .14 for the Class of '90.

This has presented us with a problem when attempting to evaluate changes in the remediation programs we have instituted by comparing successive classes, since we cannot reliably compare the preparedness of one class with another. The Elementary Algebra CPT improves our ability to compare classes, but only marginally. The new College-Level Mathematics CPT has a substantially higher correlation with first quarter calculus grades ($r = .60$) for the Class of '94. Comparison of this new class with future classes should be more accurate. The following chart shows average scores on the MSAT and CPT tests for the Classes of '90 through '94 as well as their first quarter calculus grades. The CPT was the Elementary Algebra test for the Classes of '90 through '93 and the CLM for the Class of '94. The correlation coefficients, r, are also shown.

Class	Avg MSAT	Avg CPT	Avg Calc I	Avg MSAT/CPT	r MSAT/Calc I	r CPT/Calc I
'90	594	104	1.98	.38	.53	.46
'91	599	105	1.79	.37	.37	.39
'92	604	106	2.40	.38	.40	.31
'93	605	104	2.33	.39	.20	.32
'94	597	60.7	2.10	.52	.38	.60

A new procedure was instituted with the Class of '91 in which those students identified as weak by the MSAT and the CPT were placed in separate sections (the extended sections referred to above). The decline in the correlation coefficients between the MSAT or CPT and Calc I, compared with the Class of '90, reflects, presumably, the mitigation of grade differences caused by this new procedure. In other words, this new procedure was successful in the sense that the grades of the weak students improved. (The grades of the average students tended to decline; but that is another story.)

The New Computerized Placement Test

The results of the College-Level Mathematics CPT show that a large fraction of our incoming students are poorly prepared for beginning Calculus, in spite of the MSAT scores, which would ordinarily be interpreted as good. The median raw score was 60 out of 120 on the new CPT, which covers algebra, trigonometry, and precalculus. This points up weaknesses that are not apparent on either the MSAT or the Elementary Algebra test, where scores have averaged around 600 and 105, respectively.

Figure 1 shows the distribution of CPT raw scores by 10 - point ranges. Figure 2 shows the percentages of midshipmen falling into three broad categories of CPT scores. Educational Testing Service (ETS) has assigned tentative percentiles to the raw scores, based on results from comparable pencil-and-paper tests (the College Board Mathematics Achievement Tests, Levels I and II). The median score of 60 corresponds to a percentile rank of 46%. That is, our median student is weaker than the national median. However, it should be understood that the Achievement Tests are not taken by all college-bound students, but rather by a much smaller number of students, who are likely to be better than average in mathematics. It will be interesting to see what percentile of students taking the CLM a raw score of 60 corresponds to. There are indications that, once the CLM is normed by a sample of students taking the CLM itself, the percentiles will be higher.

ETS has provided suggested cutoff scores for placement. The minimum score required to be placed in each course is indicated below. Where two courses are shown, the choice depends on the level of the courses at each school. The number of students of the Class of '94 in each CPT range is shown, as well as the percentage of those who obtained an F in Calculus I and the percentage who received lower than a C.

The majority of our students fall into the College Algebra or Intermediate Algebra category, i.e., according to the ETS criteria, they should have been placed no higher than college algebra. The Calculus failure rates of 11% and 33% and the deficiency rates of 54% and 76% reflect this inappropriate placement.

The 77 students who should have been placed into College Algebra/Precalculus or Precalculus had a failure rate of only 4%. Either our calculus course is too easy (not likely), the ETS cutoff scores are inappropriate, or else these students are able to compensate for their deficiencies enabling them to pass the course, although the large number in the D range indicates that this compensation has its limits. It is suspected that, for many, the poor performance on the CPT reflects the several years elapsed since their last math course. Post-testing a small sample of students several

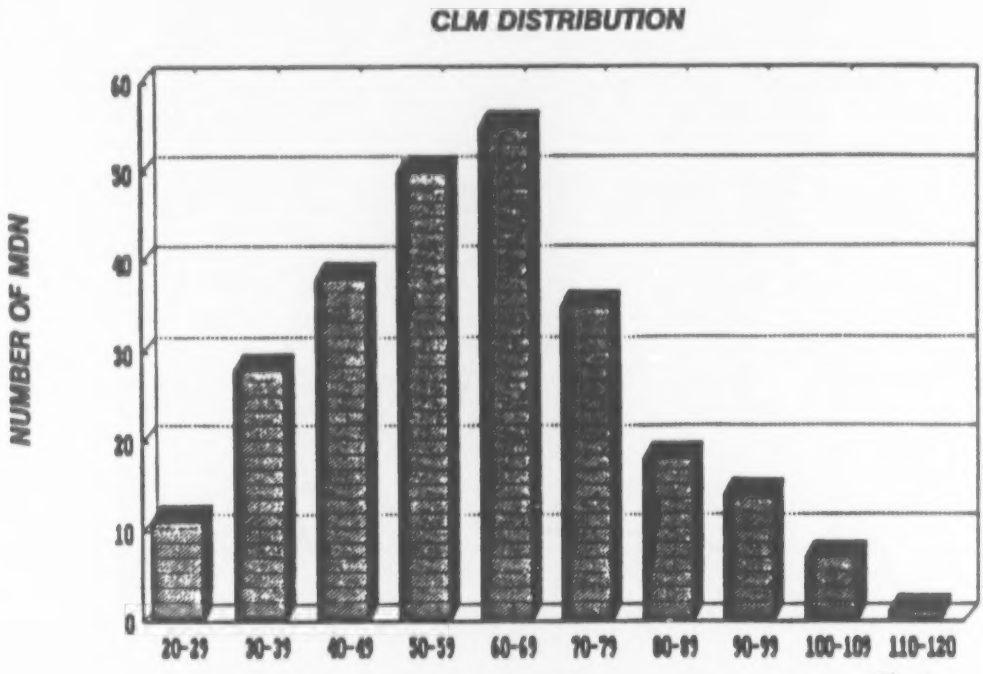

CLM DISTRIBUTION
NUMBER OF MDN
60
50
40
30
20
10
0
20-29
30-39
40-49
50-59
60-69
70-79
80-89
90-99
100-109
110-120

Fig. 1

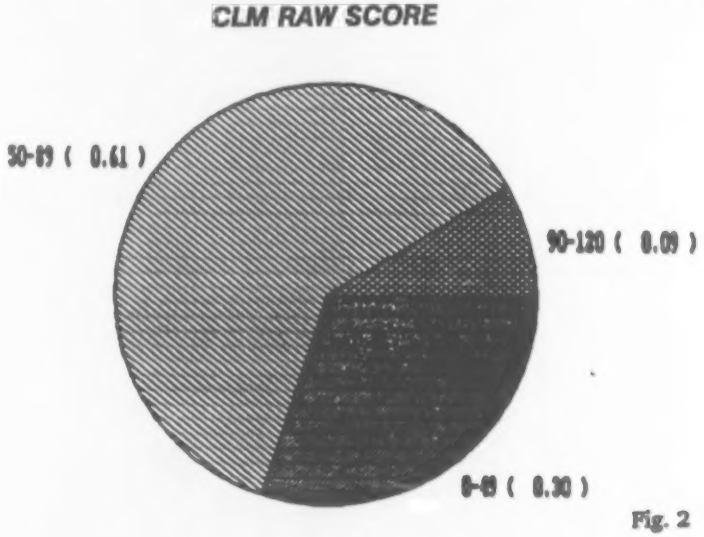

CLM RAW SCORE
50-89 (0.61)
90-120 (0.09)
0-49 (0.30)

Fig. 2

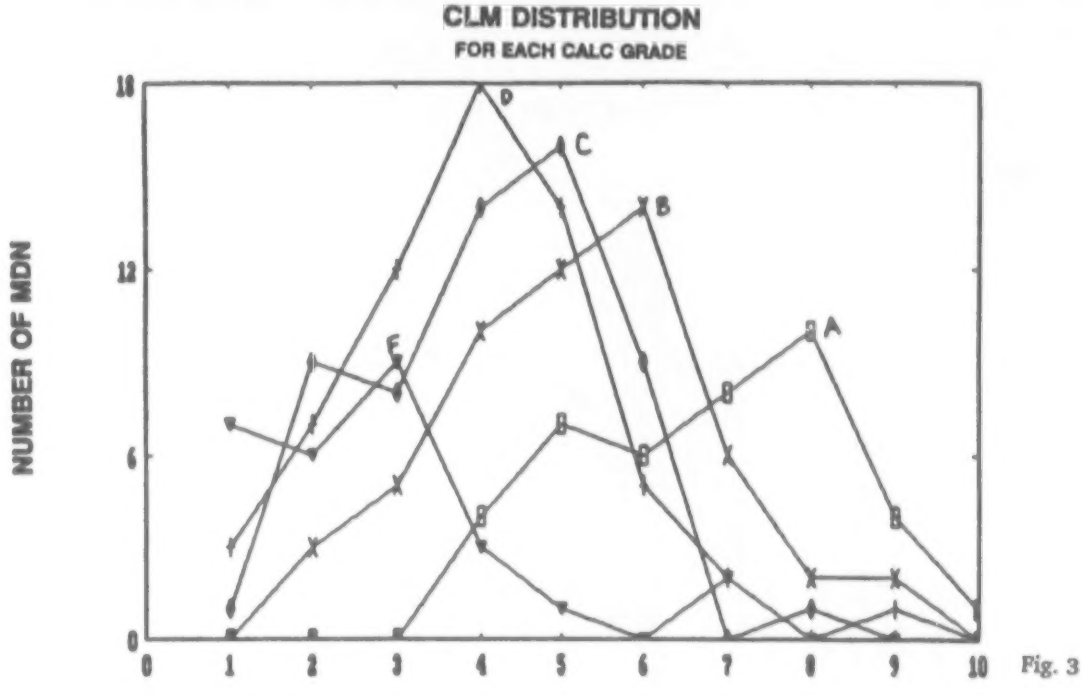

CLM DISTRIBUTION
FOR EACH CALC GRADE
NUMBER OF MDN
18
12
6
0
0 1 2 3 4 5 6 7 8 9 10
A
B
C
D
F
CLM RANGES
A B C D F

Fig. 3

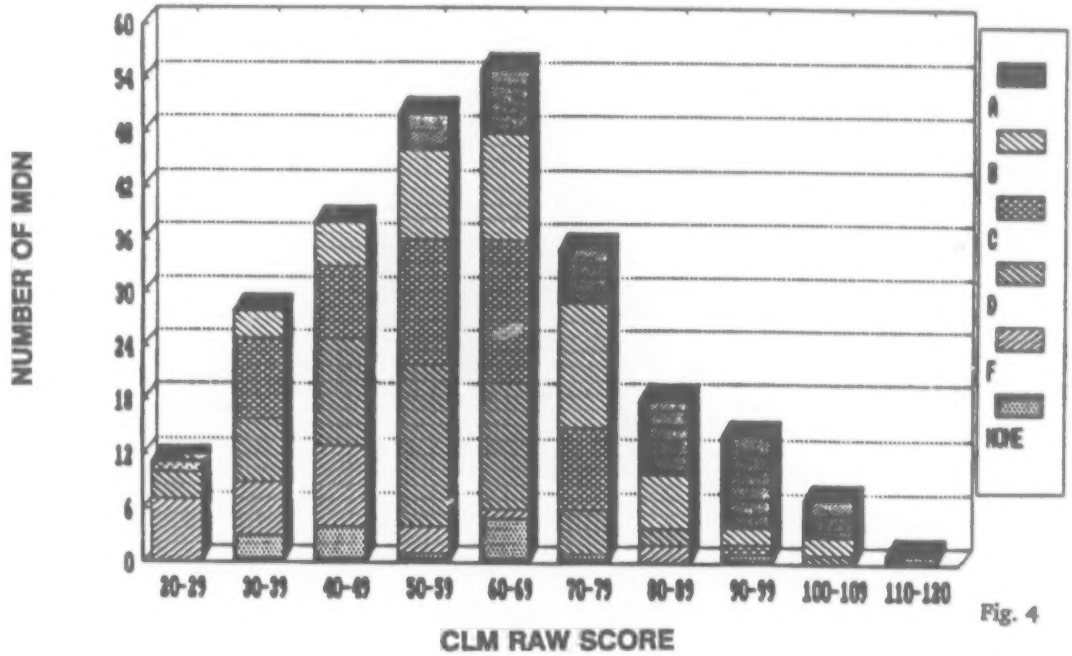

NUMBER OF MDN
60
54
48
42
36
30
24
18
12
6
0
20-29 30-39 40-49 50-59 60-69 70-79 80-89 90-99 100-109 110-120
A
B
C
D
F
NONE
CLM RAW SCORE

Fig. 4

PLACEMENT CRITERIA SUGGESTED BY ETS

CPT(CLM) Raw Score	Course	Number of Students	%F	%<C
92-120	Calculus	8	0	6
86-91	Calculus or Precalculus	11	0	9
70-85	Precalculus	46	4	22
63-69	College Algebra or Precalculus	31	4	37
43-62	College Algebra	104	11	54
<43	Intermediate Algebra	47	33	96

years ago, after 3 quarters of calculus, using the Elementary Algebra CPT, showed that all students (the survivors, of course) were able to bring their scores up to a very high level. It is clear then that, if one can survive for the first few quarters, math deficiencies, at least in the area of elementary algebra, can be effectively remedied within the normal calculus sequence. It is suspected that this is also partially true of intermediate algebra and trig. The problem of survival remains, however. Some kind of remedial program is still needed both to increase the survival rate (giving weak students time to come up to par) and to brush up on skills in intermediate algebra and trig more effectively and thoroughly than can be done in the calculus course.

We used a cutoff raw score of 46 (28% national percentile), in combination with an MSAT cutoff of 600, to determine which students to place in extended math (the regular calculus course, with an additional day per week of remedial algebra and trigonometry). (Students had to be below both cutoffs in order to be placed in extended math.) The ETS guidelines, as well as the performance of these students, suggest that the amount of remedial mathematics given in extended math is much too little. It is further suggested that many of the students who were not selected for the extended sections are in need of some kind of remedial program.

Calculus Grades

Analysis of Calculus 1 grades at the end of the first quarter of this academic year, together with the results of the CPT, clearly shows that the inadequate mathematical background of our entering students (or insufficient retention of what was learned in high school) is an important cause of poor performance in calculus. Of all students with CPT raw scores of less than 46, 32% failed (57% received D or F). The failure rate for the remainder of the students is 6% (27% received D or F). The explanation does not seem to be that the students with lower CPT scores had harder graders for teachers. Useful information for next year: The failure rate for students above a cutoff of 50 (rather than 46, and ignoring the MSAT) was only 3%; i.e., half of the failures for the students with CPT of at least 46 were in the 46 through 49 range. Also, only 6 students with a CPT score below the cutoff had an MSAT score of 600 or above, 3 of whom failed. Despite the small numbers, this suggests that perhaps the MSAT scores should be ignored in making placement decisions, relying entirely on the CPT.

CLAC I GRADE DISTRIBUTION

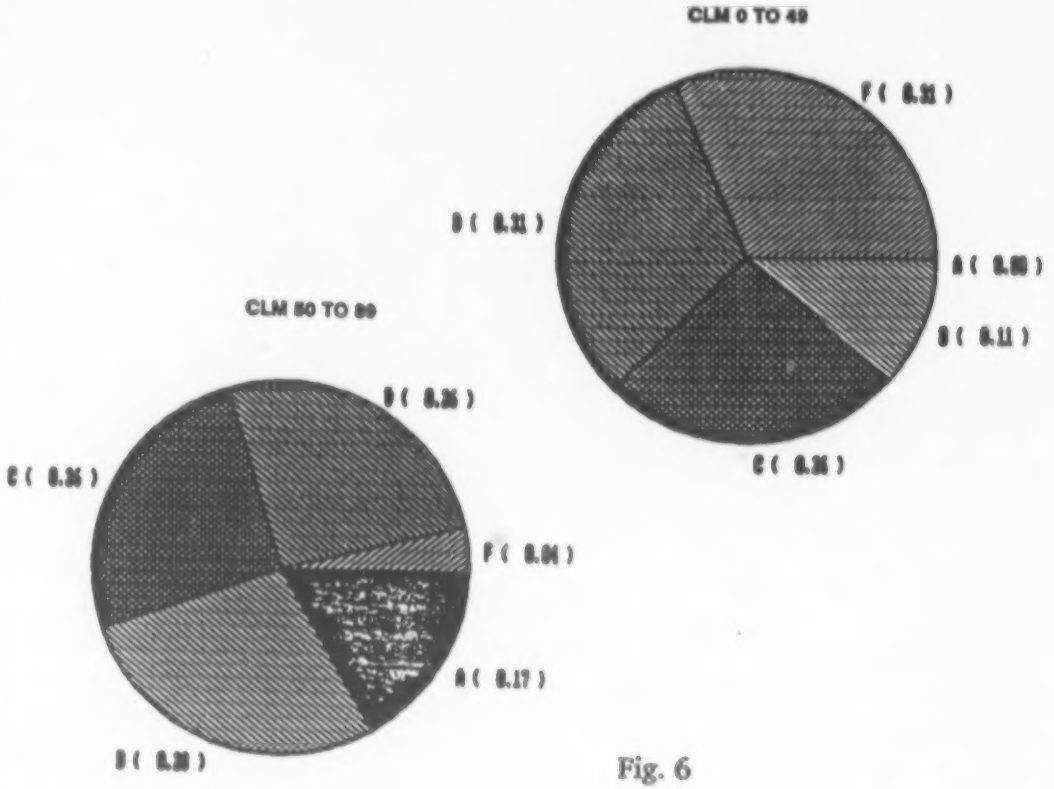


Fig. 5

Fig. 6

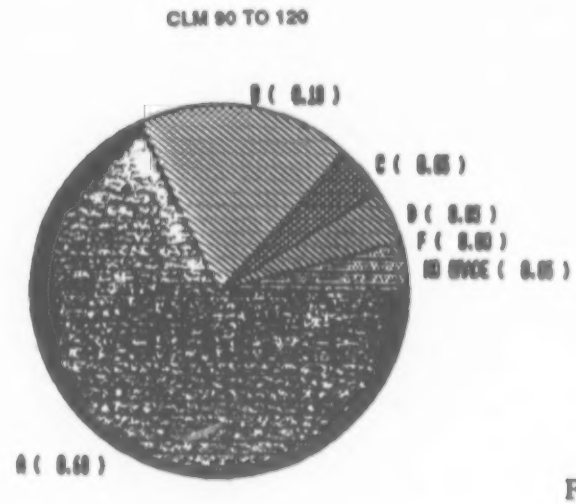


Fig. 7

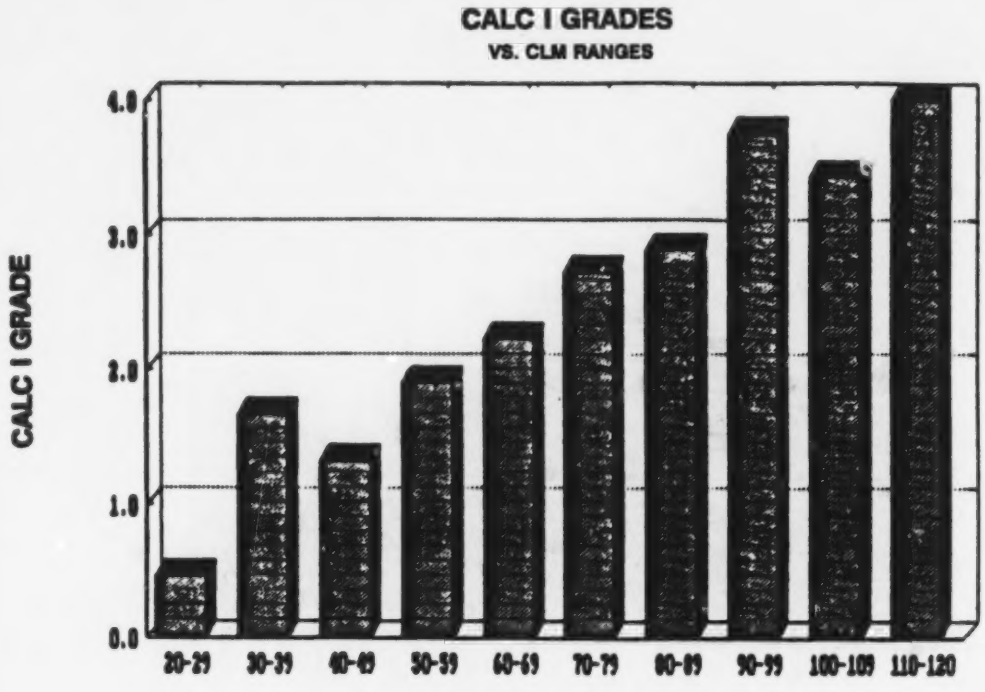

CALC I GRADES
VS. CLM RANGES
CALC I GRADE
4.0
3.0
2.0
1.0
0.0
20-29
30-39
40-49
50-59
60-69
70-79
80-89
90-99
100-109
110-120
RAW CLM SCORE

Fig. 8

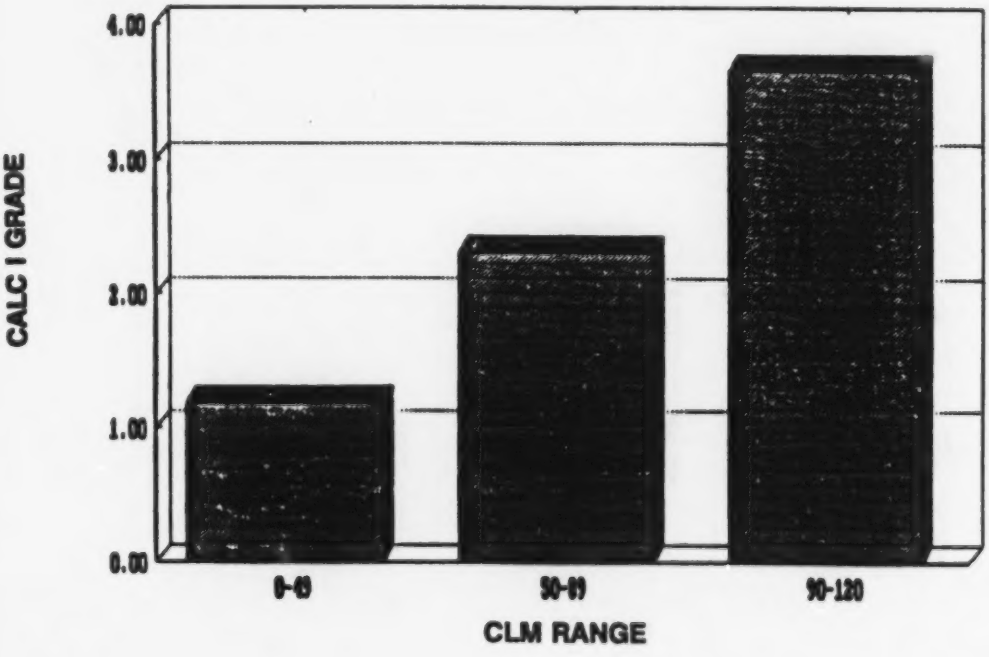

CALC I GRADE
4.00
3.00
2.00
1.00
0.00
0-49
50-89
90-120
CLM RANGE

Fig. 9

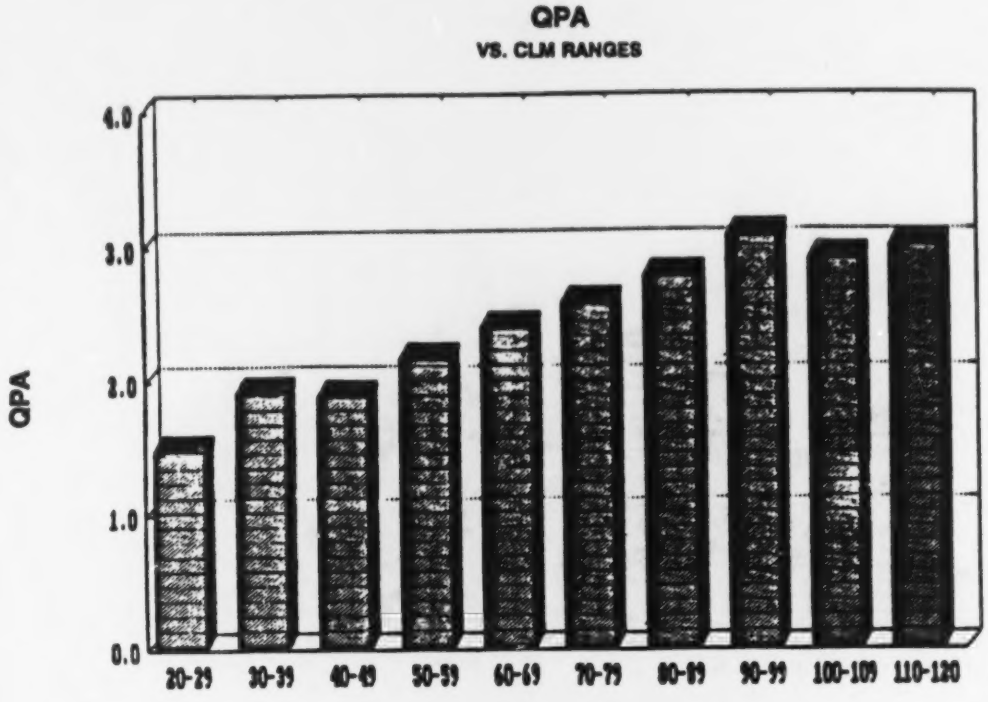

QPA
VS. CLM RANGES
QPA
4.0
3.0
2.0
1.0
0.0
20-29
30-39
40-49
50-59
60-69
70-79
80-89
90-99
100-109
110-120
RAW CLM SCORE

Fig. 10

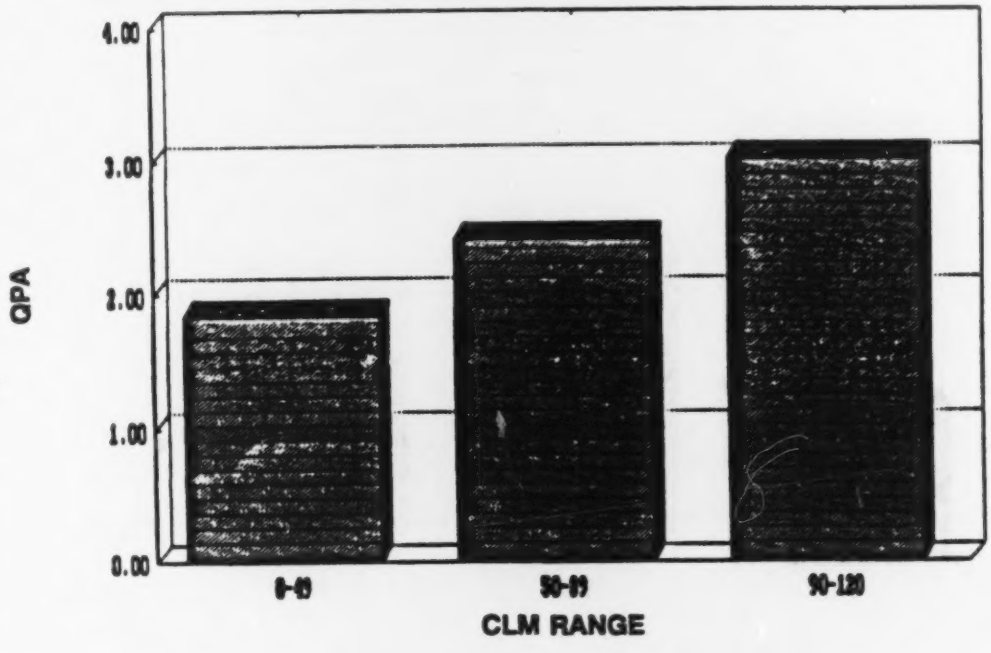

QPA
4.00
3.00
2.00
1.00
0.00
0-69
70-89
90-120
CLM RANGE

Fig. 11

Figures 3 through 9 show calculus grades vs. CPT from various points of view and in various formats. Figure 3 shows the distribution of CPT scores separately according to calculus grade. The 10 CPT ranges refer to the same 10 ranges as in Figure 1. The A students had a distribution peaking at range 8 (90-99), whereas the F students were concentrated in range 3 (40-49). While the 5 graphs are appropriately shifted relative to each other, there is still a great deal of overlap, reflecting factors other than the CPT that influence the calculus grades. Figure 4 presents the same information in the form of a stacked bar chart. The pie charts in Figures 5, 6, and 7 give the calculus grade distributions for midshipmen in each of 3 broad CPT ranges, expressed as a percentage of all midshipmen in each CPT range. The grade differences between these 3 subpopulations are striking. Average calculus grades vs. CPT are shown in Figures 8 and 9. QPA vs. CPT is shown in Figures 10 and 11 and QPA distributions vs. CPT in Figures 12, 13, and 14.

Regression Analysis

The data were analyzed in a variety of ways: overall, by instructor, by section, by company, and by floor in the barracks. Comparing all calculus grades vs. MSAT scores gave a coefficient of determination (r) of 15%, while using CPT scores as a "predictor" of calculus grades gave 37%. The multiple correlation coefficient (predicting calculus grades using both MSAT and CPT) was only slightly higher than the correlation coefficient using only the CPT as a predictor; the coefficient of determination was 38%. The CPT test was therefore a substantially better predictor than the MSAT for the Class of '94.

A regression equation was obtained for calculus grades, using the CPT as a predictor. From this regression equation, we have the following rule of thumb: each 25 points on the raw CPT score translates into one letter grade in Calculus 1. A point that might be of interest: the Calculus 1 grades were very close to being normally distributed with respect to the predicted values; or, more precisely, the residuals were normally distributed. The regression equation was used in several ways, described below.

Analysis by Instructor

It might be suspected that the reason the weak students did poorly is that they were placed into the extended sections and given instructors who grade hard. In order to determine whether this is true, the grades must be analyzed by instructor. For each instructor, the calculus grade for each student in class was predicted by the regression equation. The predicted grades were averaged for each instructor and then compared with the actual grade average for each instructor. There were 5 instructors teaching Calculus 1 in the first quarter. For the 3 instructors teaching the extended sections, the average discrepancies between the predicted and the actual grades were -.09, +.06, and +.15, where a full letter grade discrepancy would be plus or minus 1.00. Thus, those instructors teaching extended sections graded very close to the predicted grades, on average.

It is possible, however, that, since the slope of the regression line was determined in large part by the instructors with the extended sections, their grading close to the regression line was preordained. To some extent this is true. However, Figure 15 shows the regression lines determined individually for each instructor. The slopes of the regression lines varied from .027 to .049. (The overall regression line has a slope of .040.) The higher the slope, the greater the disadvantage for the more poorly prepared students relative to the stronger students. The three instructors for the extended sections had slopes of .036, .045, and .049. The slope of .049 is .009 higher than the overall slope. This represents a grade difference of less than one letter grade per 100 points on the CPT. This is insignificant. There was no instructor who could be clearly identified as grading weak students especially hard.

Nor could any of the extended section instructors be identified as grading hard for all students, strong as well as weak. Only one of the regression lines lay uniformly below all the others, but this instructor did not teach any extended section.

To further determine the extent to which the grades of the students in the extended sections were influencing the regression line, a recalcula-

QPA DISTRIBUTION

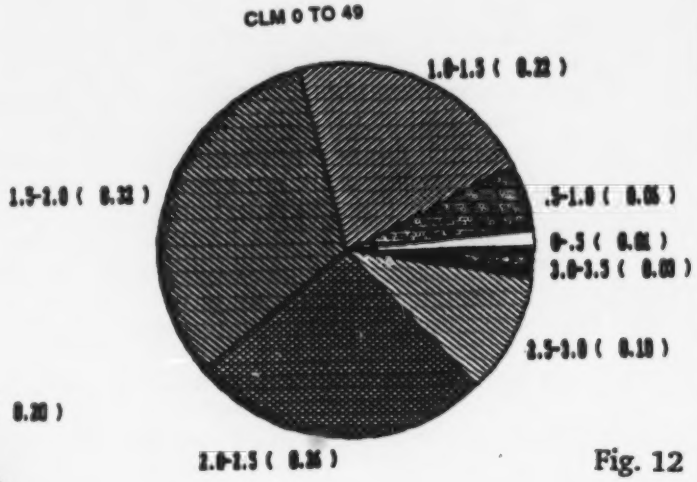


Fig. 12

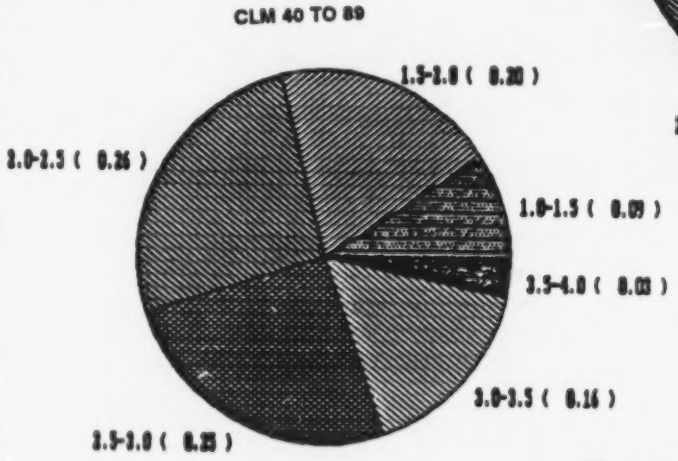


Fig. 13

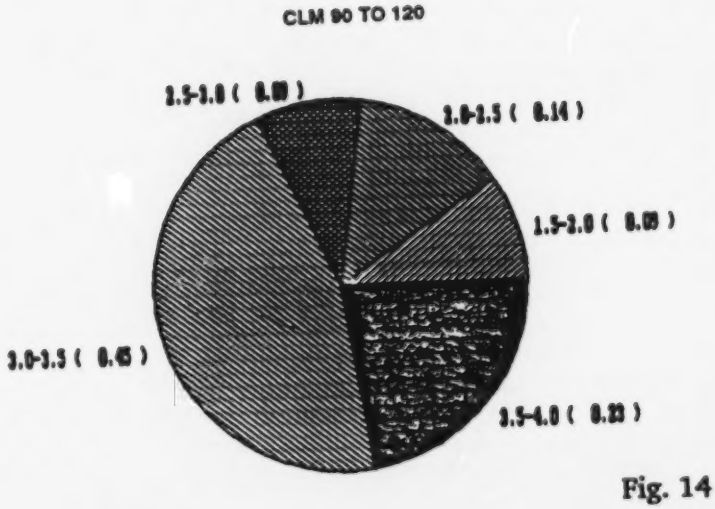


Fig. 14

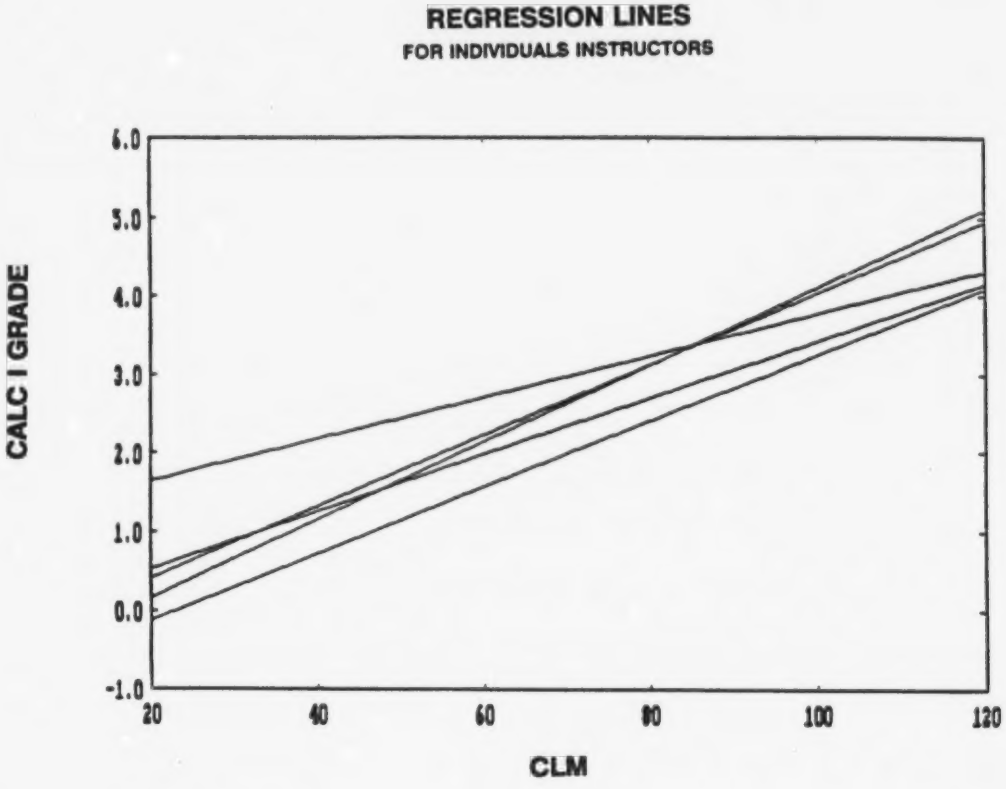

REGRESSION LINES
FOR INDIVIDUALS INSTRUCTORS
CALC I GRADE
6.0
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
20
40
60
80
100
120
CLM

Fig. 15

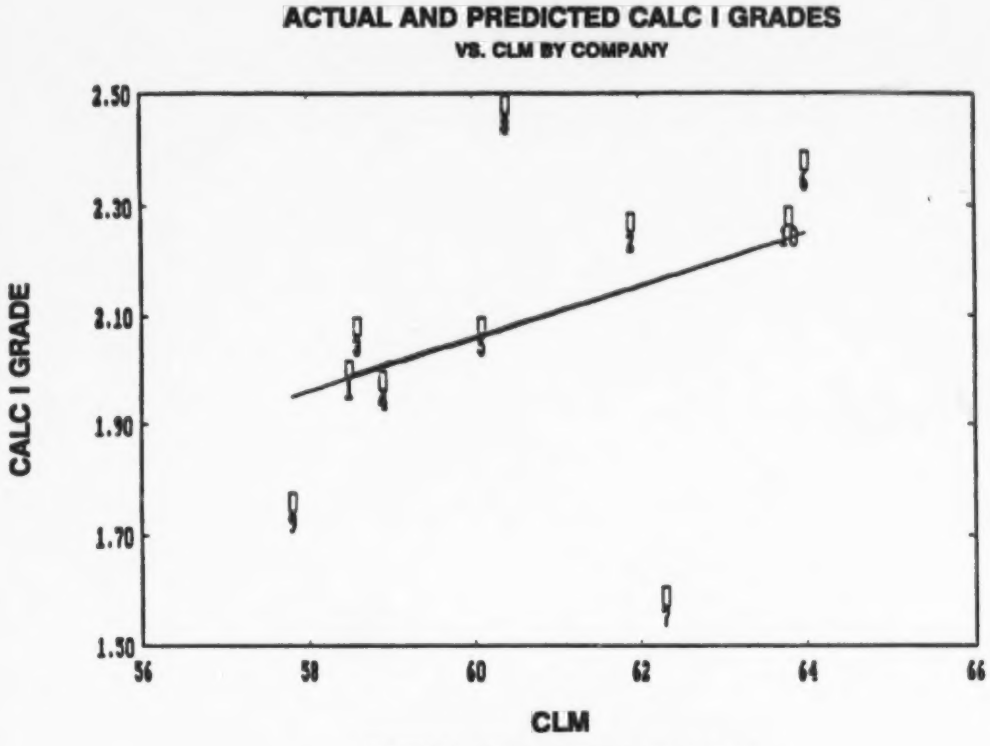


Fig. 16

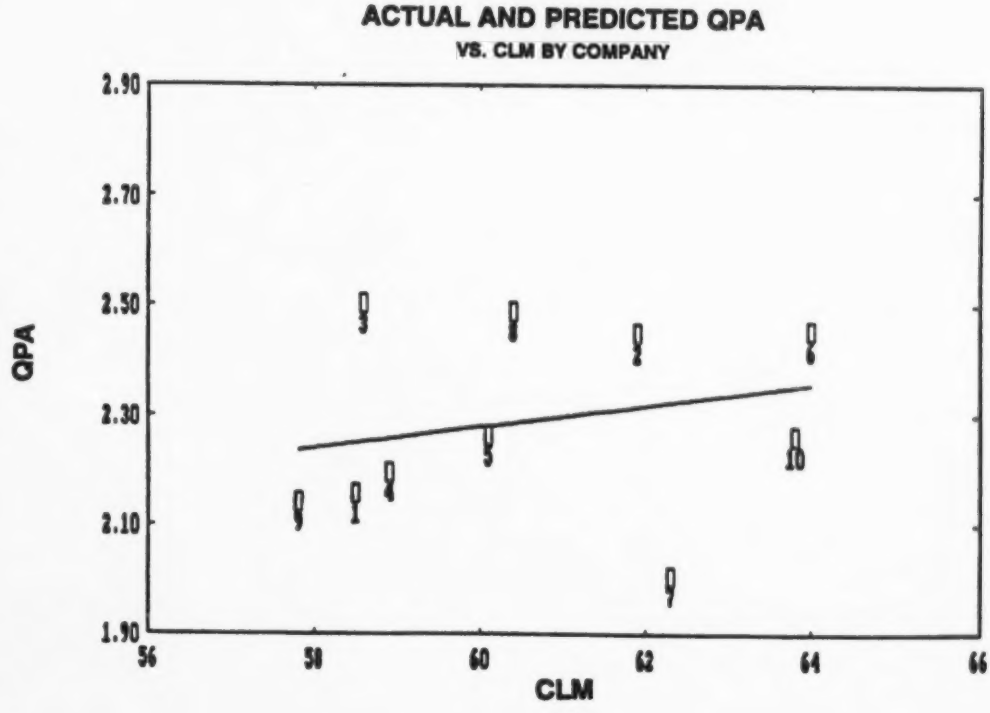


Fig. 17

100

tion of the regression equation was done, using only the regular section grades. The resulting regression equation was virtually identical to that obtained using all students. The students in the extended sections tended to lie on the regression line (on average) rather than substantially change it.

Analysis by Company

A similar comparison of predicted vs. actual grades in calculus was done by company. This is shown in Figure 16. Seventh Company scored more than half a letter grade below expectations on average, a result that bears investigation. Adjusting the standard deviation relative to the regression line for the "sample" size (company size), we find that the deviation of the calculus average for Seventh Company lies more than 2.5 standard deviations below the regression line. This indicates its high statistical significance; in other words, this is not a statistical fluctuation. Eighth Company was more than 2 standard deviations above the regression line, another significant result. Ninth Company was about 1 standard deviation below the line, a case of borderline significance; there is some possibility that this is a statistical fluctuation, but it is likely to be real.

A similar analysis by company was done for QPA. Figure 17 shows the company QPA averages vs. CPT. Significant deviations from the regression line are Third and Eighth Companies at almost 2 standard deviations above the regression line and Seventh Company at 2.5 standard deviations below the regression line. The deviations of Companies 2, 9, and 10 have borderline significance.

Analysis by Floor in the Barracks

A similar analysis was done by barracks building and floor within each building. (The first digit is the building number and the second is related to the floor, with the lower number corresponding to the ground floor.) Students on the lower floor for each building received grades that were, on average, .09 below expectations and those on the higher floors received grades .12 above expectations. This discrepancy held for all buildings except building 3, in which those on the lower floor did slightly better than those on the higher floor. The maximum discrepancy between floors was in building 2, in which there was a half letter grade (.5) difference in calculus performance relative to expectations. On average, the discrepancy was .23, or a quarter of a letter grade. This works out to about 2.5 standard deviations relative to the regression line, a small, but statistically significant difference.

We might consider the possibility that the poor performance of Seventh Company is related to the floor they reside on. Seventh and Eighth Companies share building 4 and are intermixed on the 42 and 43 floors. Average calculus grades relative to expectations for students in Seventh and Eighth Companies on floors 42 and 43 are shown below.

	7	8
43	-.34	+.53
42	-1.22	+.34

It is clear that the problem is primarily related to company. The worst performance was by Seventh Company plebes on the lower floor, where calculus grades were more than a full letter grade below expectations. There were only eight students in this category and there might be some special coincidence relating them. However, they were in seven different calculus sections.

Calculus Grades and QPA by Section

Figure 18 shows the average Calculus 1 grades by math section, along with the overall regression line. Notice the poor performance of the "extended" (B) sections on the left side of the graph. This is to be expected, since these are the weakest students. Separating these students from the other sections and giving them a remedial program is not going to turn them into strong or even average students. Our hope is to keep them from failing. This program is successful to a point. The

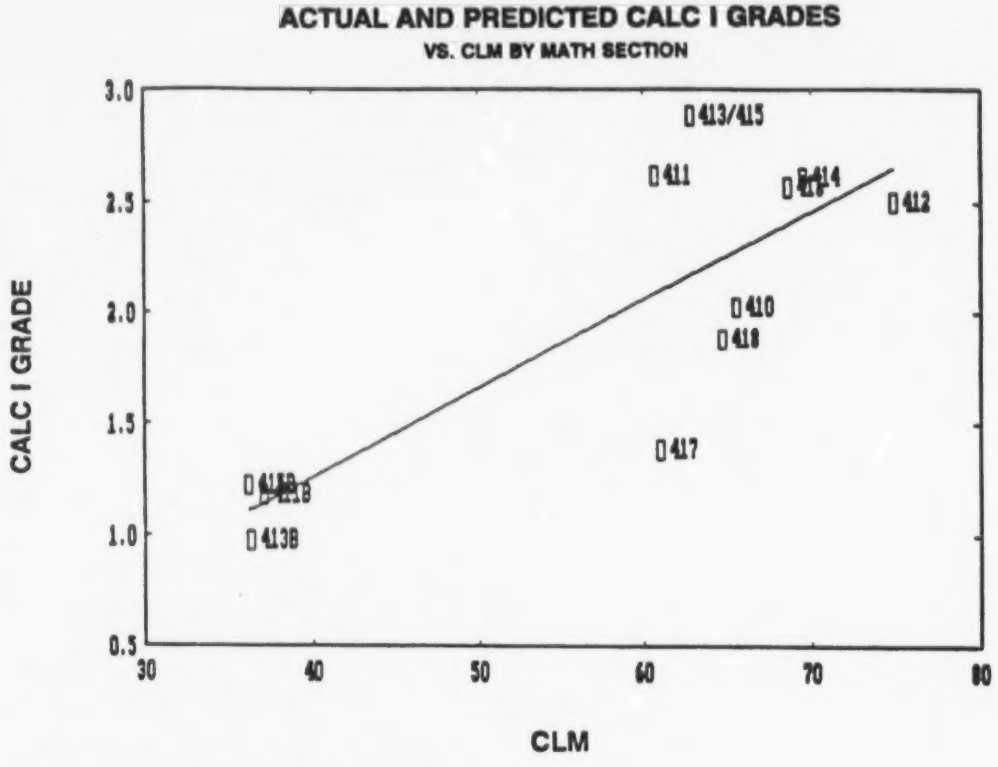

ACTUAL AND PREDICTED CALC I GRADES
VS. CLM BY MATH SECTION
CALC I GRADE
CLM
3.0
2.5
2.0
1.5
1.0
0.5
30
40
50
60
70
80
413/415
411
414
412
410
418
417
413B

Fig. 18

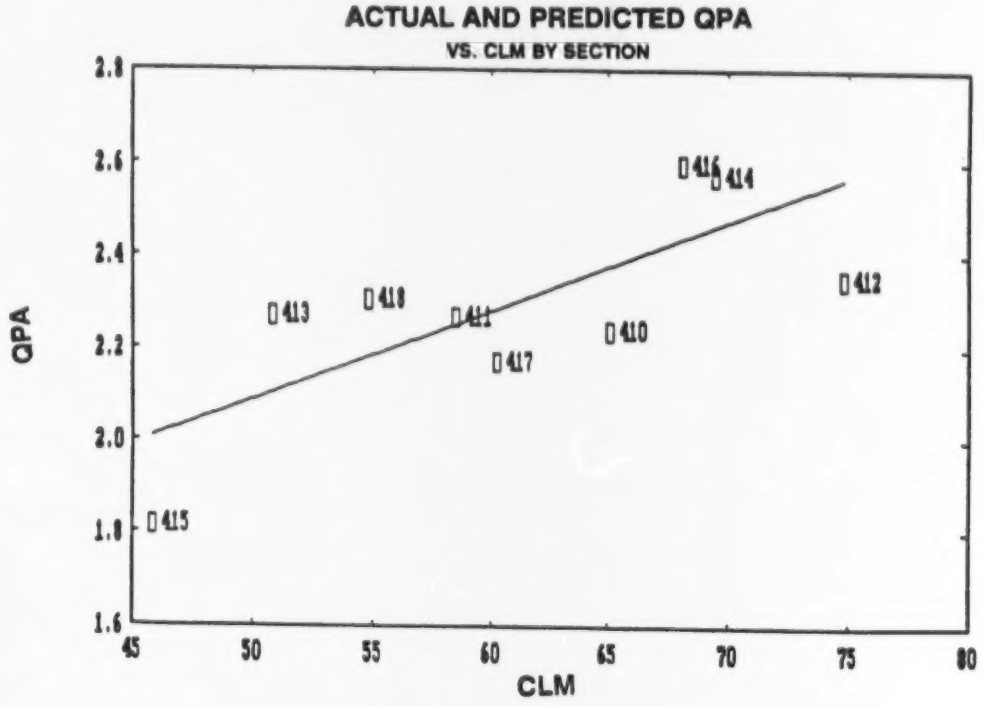

ACTUAL AND PREDICTED QPA
VS. CLM BY SECTION
QPA
CLM
2.8
2.6
2.4
2.2
2.0
1.8
1.6
45
50
55
60
65
70
75
80
41.6
41.4
41.3
41.8
41.1
41.0
41.2
41.7
41.5

Fig. 19

math faculty is looking into alternative ways to remedy math weaknesses.

It is obvious that there is a considerable amount of scatter in these data. It is important at this point to realize that the regression line in the graph is not the regression line of these 11 data points, but rather of the 250 or so data points representing all the freshmen taking calculus. The large deviations from the regression line in these samples (each section can be thought of as a sample from the population of all calculus students) shows that these are not random samples. Correcting for sample size, deviations from the regression line are as large as 4 standard deviations (sections 417, and 413/415). The deviations of sections 410, 411, 413/415, 417, and 418 are all greater than one standard deviation. These deviations are all real – they are not statistical variations. The magnitudes of the deviations range from .7 letter grade above the line to .7 letter grade below the line. Looking at different sections taught by the same instructor, we can see that part of this difference in performance between the sections is due to different grading practices of the instructors, but there is also significant variation between sections for a given instructor. For example, while section 417 was positioned .7 below the regression line, two other sections taught by the same instructor (410 and 418) were only .3 and .4 below.

As expected, the CPT is not as good a predictor for QPA as it is for calculus grades. (The coefficient of determination for QPA is 27%, compared with 37% for calculus.) However, it does help to explain the poor performance of section 415. The average QPA for each section at the end of the first quarter is shown in Figure 19.

Section 415 performed the poorest, by far, of the nine sections. However, their CPT scores were the lowest too. No section seems to have performed that badly in terms of what was predicted by the CPT. The maximum discrepancy from the regression line is well less than one-third letter grade. The poor performance of section 415 (.47 below the average for the entire freshman class) is mostly accounted for by the poor CPT scores, with only .19 remaining unexplained.

The deviations from the regression line were, with the exception of section 411, within the one to two standard deviation range. Therefore they are statistically significant. The differences between sections, apart from varying mathematical ability as measured by the CPT, are real. Explanations include the mix of instructors of varying expectations, ethos of the group ("section mentality"), varying abilities in areas not tested by the CPT (innate mathematical ability, verbal ability, etc.), as well as other unidentified causes. The MSAT does not do as good a job in predicting QPA as does the CPT. In fact, the coefficient of determination is only 13%. To repeat: the MSAT was not only a poor predictor of calculus grades, but it was an even poorer predictor of first quarter QPA.

ACT Scores

ACT scores were available for about 40 students. An analysis was done using the latest available ACT scores (not necessarily the highest). The coefficient of determination of the ACT score as a predictor for the calculus grade is 44% (higher even than the 37% for the CPT). The most valuable of the subscores is the Intermediate Algebra/Coordinate Geometry score (subscore 2). Subscore 3 (Plane Geometry/Trigonometry) and the Scientific Reasoning score have low predictive value, roughly comparable to the MSAT. Subscore 1 (Pre-Algebra/Elementary Algebra) has a high predictive value, but not as high as subscore 2.

It is not clear whether the ACT is so much superior to the MSAT as a predictor for calculus performance (it is, at least, somewhat better), or whether an important factor here is that we are using the latest available scores. We can speculate then that using the most recent, rather than the highest, MSAT score might give more accurate predictions.

Demerits

A regression analysis of calculus performance vs. demerits awarded each student showed a negligible correlation.

Second and Third Quarter Performance

The MSAT continued to be a poor predictor of grades in Calculus 2 and Calculus 3; the coefficient of determination was less than 1% for both. The CPT (CLM) did significantly better at about 22% for both. This is not as good as the 37% for Calculus 1; this is expected, however, as the "effect" is further removed from the "cause." Of greater significance for Calculus 2 is the grade for Calculus 1. The coefficient of determination here is 52%. The coefficient in the prediction of Calculus 3 from Calculus 2 is 37%. The prediction of Calculus 3 from Calculus 1 was, unexplainably, less than 1%. It seems that the best single predictor is the grade for the previous quarter.

The coefficient of determination of the MSAT as a predictor for the cumulative QPA at the end of the third quarter was less than 1%. The figure for the CPT as a predictor was 24%.

Conclusions and Recommendations

1) The CPT appears to be a good predictor of success in first quarter calculus and even in predicting the QPA. In addition, it provides a useful tool to investigate performance irregularities by section, instructor, company, and room location.

2) By several measures, the Class of '94 was weak. The problem with this class will tend to lessen in subsequent quarters and years due to setbacks and disenrollments and also since the students who make it through the first couple of quarters will have had a chance to come up to speed in mathematics and accommodate in other respects to life at the Academy.

3) Since the MSAT is such a poor predictor of early student performance (its predictive ability for purposes of admission or for later performance might be better, however), we might give consideration to the possibility of requiring applicants with an MSAT below a certain cutoff (550?) to take either the ACT or to come to the Academy (at their expense) to take the CPT. This latter alternative affords us the opportunity to interview the applicant, as well.

Charles F. Weber is with the Department of Mathematics and Sciences at the United States Merchant Marine Academy, Kings Point, Long Island, New York. April 1991

Aims Community College Computerized Placement Tests Pilot Study, Summary Report

by Darlene Nold and Marcia Kuechenmeister

Background

Aims Community College is a locally supported two-year college that serves a large number of communities north of Denver, CO. The main campus of Aims Community College is located in Greeley, CO, a city of approximately 60,000 people. In addition, the college operates two satellite campuses in Fort Lupton and Loveland to accommodate an increasing student population. As an open admissions college, Aims serves more than 17,000 credit-seeking students each year.

Established in 1967, Aims provides a comprehensive community college educational program. In addition to degree and certification programs, the college offers a variety of remedial, developmental, and community service programs. Although the college does not turn away students 16 years of age or older for lack of academic preparation, minimum skill levels are required for admission into basic education classes and most certification programs. Students, therefore, may be required to complete developmental or remedial courses to correct academic deficiencies before enrolling in college-level courses and programs.

Aims' College Assessment Center was established fall 1980 to fulfill the need to assess the academic skills of entering students for course placement. As the enrollment grew and an increasing number of vocationally oriented students enrolled, many students were found to lack sufficient basic academic skills. The need for some standard measure of academic ability was compounded further with the increasing numbers of returning, nontraditional students who as a group brought with them a broad range of abilities.

The purpose of the Assessment Center is to assess all new students in the areas of reading, English, and mathematics to determine appropriate course placement. Students with previous college coursework or with current ACT or SAT scores are excused from assessment under certain conditions. Approximately 3,000 students are assessed each year for initial course placement at Aims.

Annually, the college assessment tests are reviewed by appointed assessment committees. Three committees in the areas of reading, English, and mathematics are composed of faculty representatives from each college division. The responsibilities of the assessment committees include evaluating the effectiveness of current assessment tests, recommending changes in tests and testing procedures, discussing assessment issues, and disseminating information to faculty.

Purpose

After evaluating the current assessment tests during Fall Quarter 1989 the assessment committees began to review other tests. A lack of consensus regarding the paper/pencil tests available prompted the Assessment Coordinator and the committees to consider a new form of testing – computer adaptive testing. With the support of the committees and the assistance of the College Board, the assessment coordinator prepared to conduct a pilot study with the College Board's Computerized Placement Tests. The primary purpose of the pilot was to investigate the feasibility of using the Computerized Placement Tests (CPTs) for initial assessment of Aims' students. This report provides a comprehensive description of the pilot project, the results, and the process used to initiate CPTs at Aims. To deter-

mine the appropriateness of the CPTs, the Assessment Coordinator studied the following areas:

- Relationship between scores on Aims assessment tests and CPTs
- Relationship between CPTs scores and final grades
- Student attitudes toward computer testing
- Cost-benefit analysis
- Logistical considerations
- Selection of cutoff scores

Methodology

The pilot study was begun Summer Quarter 1990 on the main campus with the administration of CPTs to students in intact core academic classes. Information regarding student attitudes toward computer testing, test score comparisons, correlations of CPT scores with final course grades, and logistical issues was obtained.

During Fall Quarter 1990, new students required to take assessment tests took both the current paper/pencil tests and CPTs to further compare scores. The final phase of the project involved testing additional students in basic academic classes and working with faculty to establish cut-offs.

Sample

An unduplicated total of 545 students were tested between June 1990 and April 1991. Two hundred forty-two (242) students enrolled in 17 basic academic courses were tested using CPTs at the beginning of Summer Quarter 1990. Many students were tested in more than one class. The courses involved in Phase One of the pilot and the tests administered included (see Appendix A for course titles):

ENG 095, ENG 105, ENG 121
Reading and Sentence Skills

BUS 141, BUS 142, BUS 143
Reading and Sentence Skills

BUS 121, MGT 208, BIS 105, MAT 110
Reading and Arithmetic

RADIOLOGY
Arithmetic and Elementary Algebra

MAT 095, MAT 097, MAT 111
Arithmetic

MAT 112
Elementary Algebra

REA 015, REA 101
Reading

An additional 193 students were tested during the fall and winter quarters. Of these students, 121 were from intact classes and 72 were new, entering students. A small number of students in this group had taken other CPTs during the Summer Quarter. During the first week of Spring Quarter, 110 students enrolled in English composition and business English classes were tested.

Procedure

In the first phase of the study, instructors assigned to basic academic courses were contacted before the quarter started and were given the opportunity to participate in the CPT pilot. Those instructors who volunteered their classes for testing were assigned a time during the first two weeks of the Summer Quarter. Most instructors used class time for testing. Others encouraged their students to test after class.

Before Fall Quarter, entering students took both the current paper/pencil assessment tests followed by the corresponding CPTs. Students were told that the computer tests were a segment of the assessment process. Students were limited to three paper/pencil tests and three CPTs because of time and fatigue factors. A rest break between the two forms of testing was offered.

At the request of faculty, intact core academic classes were administered the CPTs at the beginning of each quarter. In addition, reading faculty used CPTs for pre- and post-testing.

The CPTs administered included Reading Comprehension, Sentence Skills, Arithmetic, and Elementary Algebra. The College Level Math test was not used during this pilot study. A com-

puter lab located near the Assessment Center was used to administer the tests.

Instructions to Students

As a group, students were given some general information about the tests they were taking. They were assured that they did not need experience with computers or typewriters to take the test. They were also told there was no time limit; in a few cases, however, students were not allowed to complete their testing because the lab was needed for a class. In most cases, students needed no more than 20 minutes to complete any one test; the average time per test was 15 minutes.

Students were not told it was a pilot, the testing was described as additional information for the instructors. The students were assured that the results would not affect their course grade or placement. In addition to the oral introduction, an information sheet was set at each computer for students to read. A copy of the information sheet and the instructions given to students is provided in Appendix B.

Administration of Tests

At least one staff member was in the lab at all times to answer questions and to move students past the tests they were not taking. When testing was completed, the students responded to a questionnaire to assess their attitude toward computer testing.

Scoring and Distribution of Scores

Scores and personal data were transferred by disk from the lab computers to the Assessment office computer where they were printed. CPT score reports with Aims assessment scores were sent to instructors. An explanation of the scores, statistical analyses, and samples of cutoff scores selected by two community colleges using CPTs were included to assist instructors with the interpretation of scores.

New students who took both forms of tests were given feedback on the two forms of assessment. They were told what the CPT scores meant but that cutoff scores had not yet been established for the CPTs. Students were told that placement decisions were made according to scores on the paper/pencil tests.

Questionnaire Development

The student questionnaire was adapted from a questionnaire developed by the College Board and ETS for use in earlier pilot testing. A demographic information section was added to the questionnaire to allow analysis of responses by specific variables. A copy of the questionnaire with the results is provided in Appendix C.

Faculty Involvement

Early involvement of the faculty and staff was essential to the support and success of CPT implementation. To familiarize faculty with the CPTs, several computers were set up for testing at three sites on the main campus. The faculty was encouraged to take the CPTs and make comments. Some of their concerns included student computer anxiety, test content, and score interpretation. These concerns were addressed by the coordinator and staff of the Assessment Center during the pilot study.

Results

Summary statistics and frequency tabulations were performed on the CPT scores. Score ranges and test score means of students taking the same subject tests were compared by course. For example: mean Arithmetic scores of students in MAT 095, MAT 097, MAT 110, and MAT 111 were compared. Correlations and simple regression analyses were performed. CPT scores were correlated with current assessment scores and end of quarter grades.

Responses to the questionnaire were studied using summary statistics, frequencies and cross tabulations. Chi-square was performed to identify any significant differences within the variables: age, gender, and computer experience. The results are reported on the primary interest areas studied to determine feasibility.

Relationship Between Scores on Current Assessment Tests and CPTs

To determine the relationship between the current assessment and the CPTs, simple correlations and regression analyses were performed on the raw scores of the current assessment and the total right scores of the CPT. Percentile distributions were also constructed to establish cut-offs. The scores of only those students who had taken both types of tests within three months of each other were included in the analyses.

Reading

Aims' current reading assessment test is the Diagnostic Reading Test, published by the Committee on Diagnostic Reading Tests Inc. It is a 40-minute timed test composed of three separate sections: general reading, vocabulary, and reading comprehension. The test provides a grade level equivalent with a reading rate as well as comprehension and vocabulary raw scores. The total raw score on this test was used to compare with the total right score on the CPT Reading Comprehension test. The scores of 134 students were used to make the comparisons. Correlation between the tests was performed ($r = .66$).

English

The Missouri College English Test is currently used to assess the English skills of entering students at Aims. The Missouri Test is a 90-item proficiency test that measures language mechanics and effectiveness of written expression. The 40-minute test covers punctuation, capitalization, spelling, grammar, sentence structure, and paragraph organization. The scores of students who had taken both English tests ($N = 105$) were compared. The two English tests correlated at .77.

Arithmetic

The current arithmetic assessment test used at Aims is locally developed. It is a 37-item multiple-choice test that covers addition, subtraction, multiplication, and division of whole numbers, decimals, and fractions. The test includes conversions, word problems, measurement, positive and negative numbers. Students are given 40 minutes to complete the test, and no calculators are allowed. The scores of 138 students were used for comparison. A correlation of .73 was found between the current Aims' arithmetic assessment and the CPT Arithmetic Test.

Algebra

Aims' entry algebra assessment tool, the Intermediate Algebra test, is a 38-item locally constructed test. It covers arithmetic, polynomials, factoring, fractions, equations, formulas, system of equations, word problems, and property and inequality problems. Students are allowed 50 minutes for the test, and calculators are not permitted. Very few students ($N = 14$) had recent, entry intermediate algebra scores to compare with CPT Elementary Algebra scores. Most of Aims' students progress into the intermediate algebra course from prerequisite beginning algebra. Correlation between the two tests was found to be .78 despite the low number of students.

Relationship Between CPT Total Right Scores and Course Final Grades

In addition to making direct comparisons between scores on the CPTs and Aims' current assessment tests, the researchers studied the relationship between CPT scores and final grades in basic academic courses. Most of these data were obtained from the students tested in intact classes. New students were tracked to obtain final grades in entry-level English and mathematics classes, however, because of the few number of students tested in reading classes, correlations were not performed on the reading data.

English Classes

New students who do not place directly into the entry-level college English composition course (ENG 121) are required to enroll in a remedial (ENG 095) or review course (ENG 105) before attempting college-level English material. Students enrolled in these entry-level English courses

were administered the Reading and Sentence Skills CPTs during the first two weeks of the summer 1990 quarter. Correlations were conducted to determine the relationship between CPT scores and English final grades. Regression analyses were performed to determine the CPT scores associated with final grades.

The progress of 46 students enrolled in the developmental English course (ENG 095) was tracked. A correlation of .29 was found. Regression analysis indicated that students receiving a C or better obtained a score of 63 or higher on the Sentence Skills CPT.

Twenty-five students enrolled in ENG 105 were administered the CPT. The correlation of their scores on the Sentence Skills test with their final grades was .38. A final grade of C or better was associated with a CPT score of 89 or higher.

Twenty-three students in ENG 121 were also tested using the Sentence Skills CPT. A correlation of .18 was found between score and final grade. Students receiving a grade of C or better obtained a score of 103 or higher on the Sentence Skills CPT.

Mathematics Classes

Most frequently, entering students are initially placed into one of the following courses: remedial mathematics, beginning algebra, or intermediate algebra. Students enrolled in these entry courses were administered either the Arithmetic or Elementary Algebra CPT depending on the course. Thirty students enrolled in MAT 095, Introductory Math, were given the Arithmetic CPT. A moderate correlation (r = .45) between their CPT scores and their final grades was found. Regression analysis indicated that students receiving a grade of C or better obtained a CPT score of 42 or higher.

Final grades of 20 students enrolled in MAT 097, Introduction to Beginning Algebra, were correlated with their scores on the Arithmetic CPT. A moderate relationship (r = .55) was found. Furthermore, a grade of C in MAT 097 was associated with a score of 50 on the CPT.

Forty-seven students enrolled in MAT 111, Beginning Algebra, were administered the CPT. The correlation between their scores on the Arithmetic CPT and their final grades (r = .55) was calculated. A grade of C or better was associated with a CPT score of 62 or higher.

The final grades of 40 students enrolled in MAT 112, Intermediate Algebra, were correlated with the students' scores on the Elementary Algebra CPT. A correlation of .57 was noted. Students who received a final grade of C obtained an average score of 71 on the CPT.

Student Attitudes Toward Computer Testing

Students (N = 419) who took the CPTs during the pilot study and responded to an attitude survey ranged in age from under 18 to over 50. The majority (80%) of students were age 30 or under, with most students (53%) between the ages of 18 and 24. Males and females were represented with fifty-seven percent female and forty-three percent male students. Nearly half (43%) the students had at least one to two years of computer experience whereas 19% had none. This demographic information was used to identify any significant differences among students.

Student responses were analyzed by the primary factors of interest: test difficulty, ease of use of the computer, test anxiety, and test administration preference (paper/pencil or computer). A complete report on student attitudes was prepared separately. The following summarizes the results.

Test Difficulty

Students, overall, felt that the computer tests provided a reasonable degree of difficulty and challenge for them. In each subject area, the majority of students rated the tests about right in difficulty: Reading Comprehension (69%), Sentence Skills (60%), Arithmetic (55%), and Elementary Algebra (57%). The majority (59%) of the students also felt some challenge to do their best whereas 38% felt greatly challenged.

Ease of Use of the Computer

Although challenging, the computerized tests were rated as easy to take by the majority (96%) of the students regardless of age, gender, or computer experience. With only simple instruction, students seldom had questions regarding recording their answers. In addition to finding the program easy to use, the majority (61%) of the students found the CPTs less tiring than the traditional paper/pencil tests. Although a lack of computer literacy was an initial concern, the pilot study demonstrated that the use of the computer did not disadvantage computer-naive students.

Test Anxiety

The process of taking a test creates some degree of anxiety in most people. More than half (53%) the students who participated in the pilot, however, indicated that they did not feel nervous while taking the CPTs. Furthermore, nearly half (45%) the students indicated that they were more nervous taking paper/pencil tests than they were taking the computerized tests. The untimed feature of the CPTs appealed to students, and the majority (59%) of the students indicated they felt no pressure to complete the test quickly. The fact that the test is self-paced is expected to contribute greatly to the reduction of test anxiety.

Test Administration Preference

The majority (67%) of students in the pilot study indicated that they preferred taking the computer-administered tests over paper/pencil tests, regardless of their age, gender, or computer experience. Most of the written comments reflected students' interest in the CPTs. Students found the computerized tests were easy, enjoyable, and fun, and a good way to test. The students found them less stressful and tiring than paper/pencil tests, and students liked the fact that they were untimed.

Cost-Benefit Analysis

To determine the cost effectiveness of using CPTs, several factors were considered including

Equipment and Materials

Computer Equipment	Total Cost
10 CompuAdd computers @ $1,569 each	$15,690.00
2 Desk Jet printers @ $375 each	750.00
4 Panasonic printers @ $179 each	716.00
2 Switch boxes @ $75 each	150.00
Cables	82.26
Subtotal :	$17,388.26
Furniture	
2 Computer carts	$868.28
Tables	402.00
Chairs	270.00
Subtotal:	$1,540.28
Maintenance	
Rewire Testing Room	$244.35
Total:	**$19,172.89**
Program Fees	
CPT Software program Level I	$750.00
Background questions	500.00
Estimated number of tests per year	9,000.00
Total:	**$10,250.00**

the cost of equipment and materials, program fees, and savings on student and staff time.

Benefits

The CPT provides many advantages over the traditional paper/pencil test. Because the CPTs are administered untimed and require a minimum of instruction, assessment staff have time for more student interaction and can be more accommodating of student schedules. It also means that the staff can provide other testing, such as GED, Teacher Certification, and developmental assessment, more frequently.

Offering assessment testing on a walk-in basis will change the flow of students in and out of the center, making the flow more steady and minimizing the periodic rush of students for testing. Furthermore, during busier times, students will be accommodated more readily. This should also reduce the demands placed on the assessment staff.

Students will benefit the most from the CPT. Because the test is untimed, the pressure of a time limit is eliminated and the students are given the opportunity truly to demonstrate their skills. Because of the adaptive capability, CPTs provide an accurate indicator of ability with fewer questions than the traditional testing methods. This saves students time, gives students a clear indication of their academic skills, and assures placement into the appropriate classes. Students who are placed properly are likely to succeed and return to complete their education, which benefits both students, and the college.

In addition, this form of testing provides a positive introduction to the computer for students with no prior experience. Because students will leave the testing areas having been successful using high technology, they may also feel more confident and motivated about pursuing their academic careers.

Logistical Considerations

Implementation of computerized testing at Aims required the support and cooperation of the assessment personnel and the computer services staff. Planning was more of a challenge because all three campuses had to be prepared to initiate the CPTs at the same time.

On the main campus, six IBM CompuAdd computers and two Hewlett Packard Desk Jet printers were purchased to handle the assessment testing. Desk Jet printers were chosen because they are quiet when printing. A computer lab with 19 computers was located to accommodate the larger volume of students testing between quarters and during registration. Currently, the main campus in Greeley handles 60 percent of all assessment testing, which amounts to about 2,000 students a year.

Because placement testing is administered less frequently at the two branch campuses, less equipment was needed to prepare these sites for the CPTs. Two additional computers were purchased for each of the branch campuses, and these computers with printers were set up on carts for easy access and mobility. Computer labs with 12 to 14 computers will also be available on these campuses for testing between quarters and during registration. Fifteen percent of all students (about 450 per year) are assessed at South Campus, and 25 percent (about 750 per year) are assessed at West Campus.

Testing will be available on a walk-in basis during the quarter on the main campus where an average of 10 to 70 students a week are assessed. The branch campuses will schedule appointments because they have fewer computers and a limited number of personnel to assist with testing. Branch campuses assess an average of 2 to 30 students a week during the quarter. Again, between quarters and during registration, computer labs will be used to accommodate approximately 200 students on the main campus and from 10 to 80 students on the branch campuses.

Once approved, and before CPT testing could be initiated, a number of administrative issues had to be addressed. This included updating the minimum entry-level assessment scores provided in the Advisors' Manual and revising Student Score Interpretation forms. After the faculty was satisfied with the cutoff scores, changes in the Advisors' Manual were finalized. Student Score Interpretation forms were revised with the new

CPT scores required for placement into entry-level courses. Other administrative changes included revising the mainframe computer screens for CPT score display and preparing a program for transferring scores to the mainframe. The final activity before implementation of CPT testing involved conducting a training session for assessment coordinators and staff at the branch campuses.

Selecting Cutoff Scores

The process of selecting cutoff scores provided a challenge to faculty. To assist the faculty in the process, the Assessment Coordinator compiled information packets with the analyses of the pilot data. The packets for the reading, English, and math faculty included raw scores, summary statistics, frequency polygons, and percentile distributions. The faculty used the data to determine how parallel the two types of tests are, to view the range of scores of students enrolled in core academic courses, and to determine how well CPT scores predicted final grades. The faculty found percentile comparisons and score distributions were most useful in selecting cutoff scores.

By reviewing these analyses and studying the cutoff scores used by other colleges, the faculty was able to establish initial cutoff scores that seemed appropriate. On a quarterly and an annual basis, CPT scores will be compared with final grades to determine the effectiveness of the selected cutoff scores. Faculty will also be surveyed to obtain additional information regarding the appropriateness of placement.

Conclusions

Aims Community College chose to implement the Computerized Placement Tests in June 1991, after a thorough evaluation of this new form of testing. During the year that the pilot was conducted, faculty and staff compared test scores, studied the CPT-final grade relationship, surveyed student attitudes, analyzed the cost, considered the logistical implications, and selected initial cutoff scores. The results of the pilot demonstrated that the CPTs would be an effective and efficient method for assessing and placing Aims' students into entry-level college courses.

The comparisons of scores within each subject area across tests provided moderate correlations and indicated similarities in test content. In the areas of English, arithmetic, and algebra, the comparisons were the greatest. The relationship between both forms of the arithmetic and algebra tests is not surprising because the content is very similar, and math skills are acquired—and therefore measured—in a sequential manner. The correlation between scores on the English tests was somewhat unexpected because the types of questions on the two tests appear to emphasize different skill areas. The primary emphasis of the Sentence Skills Test is language expression, whereas the focus of the Missouri English Test tends toward language mechanics. The moderate correlation not only indicates the relationship between test scores, it also supports the relationship between language mechanics and expression.

The lowest score correlation was found between the reading tests, and this can be attributed to the obvious differences in the two tests. The CPT measures comprehension and interpretation whereas the Diagnostic Reading Test also measures vocabulary and reading speed. The lower correlation can also be attributed to the speededness of the Diagnostic Reading Test. Overall, factors influencing the correlations may have been the differences in the timing, test content, format, theoretical basis, and students' perceived importance of the testing. Regardless of these factors, the correlations indicate a reasonable degree of similarity between the two forms of tests.

The correlations between CPT scores and course final grades were low for the students enrolled in English courses; moderate correlations were found for students in math courses. The low correlations in English classes may be more indicative of differences between test and course content than between the two English tests. The higher correlations in the math areas provide additional support for the similarities between the math tests and the systematic way of teaching math concepts.

These correlations must be interpreted carefully because students were placed based on another test score or previous coursework. Furthermore,

placement scores and final grades provide measures of different achievement. The placement test is a general measure of achievement whereas a final grade is a measure of specific achievement and is influenced by factors in the classroom. Such factors as motivation, teacher effectiveness, and attendance influence students' learning and consequently the grade. These analyses do provide an indication of placement accuracy which, however, is further supported by other analyses performed during the pilot.

Students' reactions to computer adaptive testing were considered carefully when determining the value of CPTs for assessing students at Aims. Overwhelmingly, students favored this form of testing. Aims students liked the fact that the test is untimed and takes less time to complete. They felt that the test was a fair measure of knowledge. Students felt challenged, but test anxiety and nervousness were reduced with the CPT. Such issues as computer anxiety and computer inexperience appeared to be of little or no consequence. Because students strongly favored the use of the CPTs, faculty and staff were encouraged to consider this new form of testing.

The analysis of the cost of implementing CPTs indicated that the initial cost was greater than the traditional paper/pencil test; the benefits of using CPTs outweigh the costs, however. The savings in time for both students and staff, the accuracy of measurement, the flexibility, and the appropriateness of the test content more than justify the expense differences. An informal comparison of expenses between the CPT and paper/pencil testing demonstrated a smaller disparity of cost over time.

The logistical challenges of implementing computerized testing provided the greatest hurdle in the process. The cooperation and coordination of faculty, assessment and counseling staff members, and computer service personnel was essential in making the process go smoothly. The cooperation of the faculty in becoming familiar with the test and the test scores was a primary factor in promoting campuswide interest. For acquiring the additional computer equipment and the setting up of the testing areas, use of the computer service personnel was essential. The active involvement of personnel was key to the acceptance of this new form of testing.

The cooperation of the faculty was particularly essential during the process of selecting cutoff scores. It was during this time that the involvement of the faculty, including the decision to implement CPTs, paid off. With a familiarity and an understanding of the test, faculty were better informed for making the decisions regarding cutoff scores. Cutoff scores were selected with great care and deliberation, using the data from the pilot study. The faculty was satisfied with initial cutoffs but the tracking of student progress and faculty input will determine how well cutoff scores are working.

Many of the benefits of converting to computerized adaptive testing will not be obvious immediately, but CPTs are expected to have a significant impact on the college. As an accurate indicator, a time saver, and a convenient mode of test administration, the CPTs will benefit students, faculty, and staff. Because the CPTs offer a shortened, appropriate, and accurate way of testing, the assessment process at Aims will be streamlined, and placement is expected to be improved. This efficiency in testing should reflect positively on Aims as a student-oriented institution. By providing a positive initial experience, Aims Community College will continue to give students a good start in pursuing a higher education.

Darlene M. Nold

Coordinator of the Assessment Center at Aims Community College in Greeley, Colorado. She has a Master's Degree in Psychology from the University of Northern Colorado.

Marcia Kuechenmeister

Assistant Project Director in the Assessment Center at Aims Community College. She has a Master's Degree in Educational Psychology from the University of Northern Colorado.

APPENDIX A: COURSE TITLES

BIS 105	Introduction to Computer Applications I
BUS 121	College Bookkeeping I
BUS 141	Introduction to Communications
BUS 142	Intermediate Communications
BUS 143	Advanced Communications
ENG 095	Basic Communications Skills
ENG 105	Fundamentals of Composition
ENG 121	English Composition I
MAT 095	Introduction to Mathematics
MAT 097	Introduction to Beginning Algebra
MAT 110	Applied Business Mathematics
MAT 111	Beginning Algebra
MAT 112	Intermediate Algebra
MGT 208	Small Business Management
REA 015	Developmental Reading IV
REA 101	Mastering College Reading

Appendix B: Information and Instructions

Assessment Center
Aims Community College

COMPUTERIZED TESTING
Information Sheet

The College Board Computerized Placement Tests (CPTs) measure skills in reading, English, and mathematics. The testing system is easy to use; no experience with computers is needed. The test provides accurate measurement with a small number of questions. There are no time limits, and you will only have to answer between 12 and 17 questions per test.

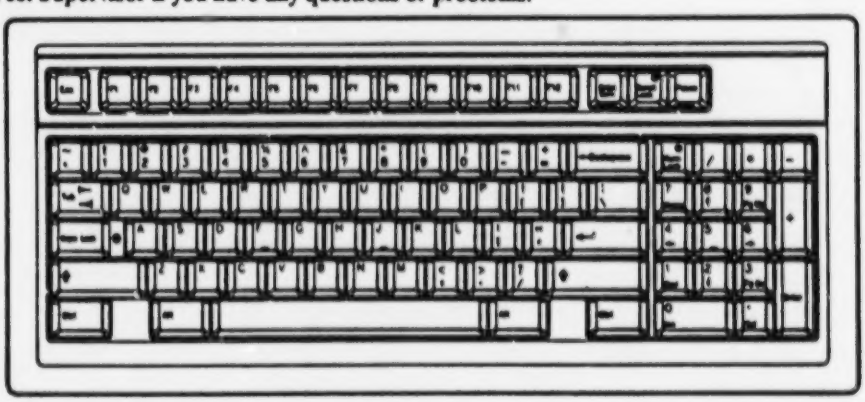
The computer selects test questions for you on the basis of your answers to previous questions. The questions are chosen to be challenging – that is, neither too easy nor too hard for you.

Because the computer uses all your previous answers in choosing the next question, you must answer every question when it is first presented. You cannot omit a question or return to change an answer to an earlier question.

The computer will explain how to record your answers as you take the test. There are three main keys that you will be using. These are the ENTER key, the F1 key (for help), and the SPACEBAR.

The first questions you will find in the adaptive test simply ask for information about you and your family – for example, how many years you have studied various subjects, or how you describe yourself.

You may begin testing when are through reading this information sheet.
Call the Test Supervisor if you have any questions or problems.



Use the SPACEBAR to move arrow to choose your answer.

Use the ENTER key to select and confirm the answer you choose.

CPTs Testing Instructions

SAY: These tests you will be taking on the computer are not pass/fail tests; they will help determine where your academic skills are now and the scores will be used to place you in appropriate classes. The tests are not timed, so you can work at your own pace.

SAY: The directions for these tests will be shown on the screen. They tell you everything you need to know and are easy to follow. You don't need to have any computer experience, and you don't need to know how to type to take these tests. Other than typing in your name and social security number, you will only need to use the SPACEBAR and ENTER keys. Notice that these keys are highlighted on the information sheet.

SAY: Before you begin the test you will answer some background questions. On the back of the information sheet is a list of major codes. You will need to use one of these numbers in response to one of the background questions. If you are undecided or are only taking a few classes, you will use 999.

SAY: During testing, the screen will blink once if you don't respond to a question after a few minutes. This is just a prompt; it isn't meant to push you. Take the time you need to answer the questions.

SAY: I will tell you which tests you will be taking. If you have any questions once testing has begun, please raise your hand and a proctor will help you.

SAY: Please put your personal belongings under your table. You should have nothing on your table except the information sheet, scratch paper, and a pencil. Aids such as dictionaries and calculators are not permitted. You will be able to pick up your profile sheet with the test scores after testing.

Appendix C: Student Questionnaire

Aims Community College
Assessment Center

STUDENT OPINION SURVEY RESULTS

Computerized Placement Tests

The Computerized Placement Test which you have just taken is a new test being used at Aims College. Because it is new for us, we would like to know what you think of this computerized test. Please answer the following questions by circling the letter of the response that is most like how you feel.

For questions 1 - 4, answer only the question/s which relates to the test/s you took.

1. If you took the READING COMPREHENSION test, how difficult was this test for you? (229 Responses)

	Percentage Response
A - Much too difficult	1%
B - Somewhat too difficult	24%
C - Just about right	69%
D - Somewhat too easy	4%
E - Much too easy	<1%

2. If you took the SENTENCE SKILLS test, how difficult was this test for you? (193 Responses)

A - Much too difficult	2%
B - Somewhat too difficult	33%
C - Just about right	60%
D - Somewhat too easy	5%
E - Much too easy	<1%

3. If you took the ARITHMETIC test, how difficult was this test for you? (203 Responses)

A - Much too difficult	3%
B - Somewhat too difficult	32%
C - Just about right	55%
D - Somewhat too easy	9%
E - Much too easy	1%

4. If you took the ELEMENTARY ALGEBRA test, how difficult was this test for you? (84 Responses)

A - Much too difficult	4%
B - Somewhat too difficult	34%
C - Just about right	57%
D - Somewhat too easy	5%
E - Much too easy	0%

5. Did you feel challenged to do as well as you could on the test? (390 Responses)

	Percentage Response
A - Not at all	3%
B - Somewhat	59%
C - Very much	38%

6. In comparison with a paper/pencil test, how challenging did you find the computerized test? (412 Responses)

A - More challenging than a paper/pencil test	28%
B - About the same	58%
C - Less challenging than a paper/pencil test	14%

7. Were you nervous while taking the test? (413 Responses)

A - Not at all	53%
B - Somewhat	42%
C - Very much	5%

8. Did nervousness while taking the test prevent you from doing your best? (405 Responses)

A - Definitely yes	4%
B - Somewhat yes	20%
C - Probably not	44%
D - Definitely not	32%

9. Which would make you more nervous, a paper/pencil test or a computer-administered test? (413 Responses)

A - Paper/pencil test	45%
B - Computer-administered test	14%
C - They'd be about the same	41%

10. Did you feel any pressure to complete the test quickly? (413 Responses)

A - Yes, a great deal of pressure	7%
B - Yes, some pressure	34%
C - No	59%

11. How easy was it for you to use the computer to take the test? (413 Responses)

A - Very easy	72%
B - Reasonably easy	24%
C - Somewhat difficult	3%
D - Very difficult	< 1%

12. Were the directions given by the computer clear and easy to follow? (414 Responses)

	Percentage Response
A - Yes, very clear	89%
B - Reasonably clear	10%
C - No, not very clear	<1%

13. Did you have any difficulty reading the material presented on the computer screen? (414 Responses)

A - Yes	11%
B - No	89%

14. In comparison with a paper/pencil test of the same length, how tiring did you find the computer test? (403 Responses)

A - More tiring than a paper/pencil test	6%
B - About the same	33%
C - Less tiring than a paper/pencil test	61%

15. Which would you prefer to take? (403 Responses)

A - A computer-administered test	67%
B - A paper/pencil test	12%
C - No preference	21%

Please give us any additional comments or suggestions about this type of testing.

_____

_____

To help us better understand your needs, please tell us a little about yourself.

My gender is: N = 172 or 43% were Male N = 232 or 57% were Female.

I am in the age range of: (404 Responses)

– Under 18	10%
– 18 - 24	53%
– 25 - 30	17%
– 31 - 35	7%
– 36 - 40	6%
– 41 - 45	4%
– 46 - 50	3%
– Over 50	1%

I have had approximately the following number of years of experience using computers (whether for professional, personal, or fun purposes): N = 404

	Percentage
– None	19%
– 1 - 2 years	43%
– 3 - 5 years	27%
– 5 - 10 years	9%
– 10 or more years	1%

I have used computers for the following purposes: (Check all that apply.)

– Word-processing
– Programming
– Data analysis
– Spread sheets
– Entertainment
– Other, please specify

Thank you for letting us know how you feel about computerized placement tests.

Validating CPT Reading Comprehension Test Standards: Employing Relevant College-Level Performance Criteria

by Anthony R. Napoli, Suffolk County Community College

Introduction

The establishment of valid placement test standards to serve as indices of academic preparedness for exemption from remedial course work and entry into traditional academic programs of study must conform to accepted and established psychometric practices. In this regard, in addition to the selection of an assessment instrument with empirically substantiated construct validity (Nunnally, 1978) and reliability (Cronbach, 1970), test users must empirically establish the *criterion-related predictive validity* (Anastasi, 1982) of the decision points (cutoffs) on a placement test's score distribution. The present report describes preliminary results of an ongoing research effort to establish empirically the cut-off points for the CPT Reading Comprehension Test (CPT-Read, 1990) at Suffolk Community College (SCC).

Based on the recommendation of the College Board/Educational Testing Service (1990) that colleges establish cutoff scores that meet the specific and individual requirements of the institution, the investigation began by selecting, with the aid of the college's instructional and administrative faculty, meaningful and appropriate criteria against which to validate CPT Reading test score cut-offs. Performance in Introductory Psychology and overall grade point average (GPA) were selected to serve as the criteria to determine college-level reading readiness. The rationale for selecting performance in the psychology course is based on the following: The psychology course, like many other core survey courses (e.g., sociology, history, economics, literature, philosophy, political science, humanities, etc.) has a heavy reliance on reading loads (typically 650 pages per semester) to cover its broad content areas. Performance in the course is dependent, to a great extent, on the adequate reading comprehension abilities typically required of college-level course work. Success in the course, therefore, serves as an index of reading skills proficiency commensurate to the demands of college curricula. From a practical perspective, the high enrollment associated with the course would substantially guarantee (far beyond any other survey course offered at SCC) sufficient data for quantitative analysis.

The inclusion of GPA as a validating criterion is based on its overall importance with regard to program completion and its ability to convey, in a global manner, success with college-level academic demands.

It was expected that both performance in the psychology course and overall GPA would be significantly related to CPT-Read scores and that, if sufficient correlation between the measures exists, regression analysis would identify those points on the CPT-Read distribution that predict academic outcomes with a high level of statistical accuracy. Once identified, those points would serve as the cut-off by which students would be placed into or exempted from a remedial reading course.

Procedures
Sample

Computer-based academic records for approximately 16,000 students attending classes at SCC during the years 1988 through 1990 were electronically searched to identify those individuals who had CPT-Read scores and who had not been placed into a remedial reading course. Of those, 1,450 students met the dual criteria and were selected to serve as the sample for the current study. It is important to note that although SCC has been employing a CPT-Read cutoff score of 90 and above to exempt students from taking remedial reading courses, 50 percent of the sample (or 722 students) had CPT-Read scores below 90. As can be seen in Table 1, the mean CPT-Read score for the sample equaled 86.12, the standard deviation equaled 16.27, and the range of scores equaled 27 to 120. An analysis of the form of the CPT-Read distribution failed to detect any significant deviation from normality.

Results

Results for Introductory Psychology

Pearson's Product-Moment correlation coefficient was employed to assess the relationship between the predictor variable and targeted criteria variables. As seen in Table 2, CPT-Read scores were found to be reliably related to psychology grades (PSYGPA; $r = .519$, $p. < .0001$). Results for the simple regression of PSYGPA on CPT-Read are presented in Table 3. Employing the general linear regression equation ($Y' = A + BX$), the next phase in analysis applied the regression constants (see Table 3) to the raw CPT-Read scores to produce predicted PSYGPA scores. This was done for each case in the sample. To determine the degree of concordance between predicted psychology grades and actual grades, both score distributions were dichotomized to form groups of students who scored grades of C and above and grades below the C level. Presented in Table 4 is the cross tabulation of predicted and actual grade dichotomies. Results indicate, for the 281 students who received grades below C, that approximately 76 percent were correctly classified based on their CPT-Read scores. Slightly more than 77 % of the 1,169 students who received "C" or better grades were correctly identified. Overall, a 77% concordance rate was observed between actual and predicted psychology grade outcomes. Statistically predicting 1,119 out of 1,450 outcomes is a highly significant finding ($p. < .000001$, Fisher's Exact Test).

To locate the point, or cut-off, on the CPT-Read distribution that is predictive of the C or better, or below C outcome in the psychology course, we employed the general linear regression equation. Listed in Table 5 are select CPT-Read scores and corresponding predicted psychology grades. The regression equation identifies a CPT-Read score of 75 and above as the point on the CPT-Read distribution that is predictive of the C and better psychology grades. CPT-Read scores below 75 predict below C performances in the psychology course.

Results for Overall Grade Point Average

Appearing in Table 2 are the correlation coefficients for overall GPA with CPT-Read and total credits completed (Credits). As expected, CPT-Read scores are positively and significantly related to GPA ($r = .414$, $p. < .0001$). In addition, GPA is also significantly dependent ($r = .270$, $p. < .0001$) on Credits. Regarding this relationship, as the number of credits increases so does GPA. Therefore, to control for confounding influence of Credits on overall GPA, hierarchical regression analysis (Tabachnick & Fidell, 1983) was used. In the hierarchical regression model, the variable Credits is entered on the first step of the analysis. This procedure produces residualized GPA scores which are adjusted for total credits completed. On a second step of the regression model, the residualized GPA scores were regressed on CPT-Read to produce the adjusted regression constants (regression coefficient and intercept) appearing in Table 6. In the manner described above to determine the degree of concordance between predicted GPA and actual GPA, both score distributions were dichotomized to form two groups of students, those who scored grades of C and above and those scoring below the C level. Presented in Table 7 is the cross tabulation of predicted and actual grade dichotomies. As seen in the table, of the 157 students who actually received grades below C, approximately

56 percent were correctly classified based on their CPT-Read scores. Slightly more than 70 percent of the 1,293 students who received C or better grades were correctly identified. Overall, the 69 percent concordance rate observed between actual and predicted grade outcomes is a highly significant finding (p. < .000001, Fisher's Exact Test).

Appearing in Table 5 are select CPT-Read scores and corresponding predicted GPA scores obtained from the regression model. The model indicates that a CPT-Read score of 70 and above is predictive of C and above grade point averages. CPT-Read scores below 70 predict GPA performances below C.

Summary

The objective of the current research effort was empirically to identify CPT-Reading Comprehension test standards to serve as decision points for placement into or exemption from remedial reading courses. The study employed criterion-referenced performance outcomes, specifically, performance in introductory psychology classes and overall grade point averages, to serve as indices of reading comprehension proficiency. On confirming the presence of significant correlation between the predictor and criterion outcome variables, regression analysis and statistical modeling were employed to predict performance outcomes from CPT-Reading Comprehension test scores. Actual and predicted grade distributions were then dichotomized to produce performance outcome categories of At/Above C Grade Levels and Below C Grade Levels to assess the agreement rates between predicted and actual grades. A comparison of predicted and actual performance categories demonstrates significantly high concordance rates of 77 percent and 69 percent for psychology and overall GPA, respectively. It is important to note that results for both sets of analyses suggest that, and converge on, a CPT-Read score in the low to mid 70s would serve as a reliable cutoff point for the SCC population.

In the present study, the application of statistical modeling and verification techniques produced significant and meaningful enhancements to placement decision-making. By selecting criterion-based performance outcomes, it became possible to identify CPT-Reading Comprehension placement test standards that convey performance capabilities with a high level of statistical accuracy.

As a final consideration, future assessment efforts will require continued monitoring of the students' academic performance and progress to ensure that the adopted cutoff points continue to serve as valid indices of academic preparedness.

Table 1:
Descriptive Statistics

Variable	Mean	Std. Dev.	Range		
CPT-READ	86.12	16.27	27.0	to	120.0
PSYGPA	2.50	1.08	0.0	to	4.0
GPA	2.76	.71	0.0	to	4.0
Credits	32.89	18.533	.0	to	81.0

N = 1450

Table 2:
Correlation Coefficients

	PSYGPA	GPA
CPT-READ	.519	.414
Credits		.270

Note: All correlation coefficients are significant at the .0001 level, df = 1449.

Table 3:
Regression of PSYGPA (Psychology) on CPT-Read

Dependent Variable: PSYGPA

Source	DF	SS	MS	F	p.
CPT-READ	1.0	458.11	458.11	532.79	.0001
Error	1448.0	1245.05	.86		
Total	1449.0	1703.16			

R-Square = .269
Unstandardized Regression Coefficient (B) = .0346
Intercept (A) = -.4879

Table 4:
Crosstabulation of Actual by Predicted Psychology Grades

	Predicted Grades		
Actual Grades	Below "C" Grade n (row %)	"C" Grade or Above n (row %)	Row Totals n (row %)
Below "C" Grade	214 (76.2%)	67 (23.8%)	281 (19.4%)
"C" Grade or Above	264 (22.6%)	905 (77.4%)	1169 (80.6%)
Column Totals	478 (33.0%)	972 (67.0%)	1450

Total Cases Correctly Classified = 77.2%.
Fisher's Exact Test, p. < .000001.

Table 5:
Predicted Psychology Grades and Overall Grade Point Averages for Selected CPT-Reading Comprehension Scores

CPT-Read	Predicted Psychology Grade (SCC Letter Grade)	Predicted GPA (SCC Letter Grade)
40	.90 (F)	1.52(D+)
45	1.07 (D)	1.61(D+)
50	1.24 D)	1.70(D+)
55	1.42 (D)	1.80(D+)
60	1.59 (D)	1.89(D+)
65	1.67(D+)	1.98(D+)
70	1.93(D)	2.08 (C)
75	2.11(C)	2.17 (C)
80	2.28 (C)	2.26 (C)
85	2.45 (C)	2.35 (C)
90	2.63(C+)	2.45 (C)
95	2.80(C+)	2.63(C+)
105	3.15(B)	2.73(C+)
110	3.21(B)	2.82(C+)
115	3.50(B+)	2.91(C+)

Table 6:
Multiple Regression of GPA on Credits & CPT-Read

Dependent Variable: GPA

Source	DF	SS	MS	F	p.
Covariate					
Credits	1.0	55.38	55.38	142.95	0.0001
Predictor					
CPT-READ	1.0	132.09	132.09	340.94	0.0001
Error	1447.0	560.59	0.39		
Total	1449.0	748.06			

R-Square = .251
Unstandardized Regression Coefficient (B) = .0185
Intercept (A) = .7769

Table 7:
Crosstabulation of Actual GPA by Predicted GPA

	Predicted Grades		
Actual Grades	Below "C" Grade n (row %)	"C" Grade or Above n (row %)	Row Total n (row %)
Below "C" Grade	88 (56.1%)	69 (43.9%)	157 (10.8%)
"C" Grade or Above	377 (29.2%)	916 (70.8%)	1293 (89.2%)
Column Totals	465 (32.1%)	985 (67.9%)	1450

Total Correctly Classified = 69.2%
Fisher's Exact Test, p. < .000001

References

Anastasi, A. *Psychological Testing* (5th ed.). New York: Maxmillan, 1982.

College Entrance Examination Board and Educational Testing Service. *Coordinator's Guide for Computerized Placement Tests Version 3.0.* Princeton, 1990.

Cronbach, L. *Essentials of Psychological Testing* (3rd ed.). New York: Harper and Row, 1970.

Nunnally, J. *Psychometric Theory* (2nd ed.). New York: McGraw-Hill, 1978.

Tabachnick, B. G., & Fidell, L. S. *Using Multivariate Statistics.* New York: Harper and Row, 1983.

Anthony Napoli is currently the Director of Institutional Services and Research and Assistant Professor of Psychology and Statistics at Suffolk Community College, Long Island, New York.

Correlations between CPTs Scores and Course Grades

By Mark Pomplun

Effective instruction and sucess in course work depends in part on the facility to properly place the student. The measurement tool used to place students must be a dependable and relevant instrument. One method of validating the instrument used for placement is to correlate test scores to course grades. This predictive validity coefficient, however, may yield a result lower than expected.

There are several factors that depress the correlations between test scores and course grades. Course grades and tests seldom measure exactly the same thing. The Computerized Placement Tests (CPTs) are measures of academic achievement. In placement applications, they are measures of the prerequisite skills necessary for success in a specific course. Course grades frequently measure factors other than just achievement, such as motivation, homework, and attendance. These other factors dampen the correlations between course grades and achievement tests. In addition, two other factors that lower correlations between course grades and tests are reliability and range restriction.

The reliability of course grades establishes a ceiling on how high any predictor test can correlate with them. Hence, the reliability of course grades places a limit on the possible correlation between the CPTs and course grades.

Unfortunately, because each student usually has only one grade per course, the reliability of course grades can't easily be calculated. However, in order to show how the reliability of course grades limits the correlation between the placement test and the course grade, an estimate of course grade reliability obtained during a recent study at ETS will be used. (It is an estimate because some of the assumptions that were used to calculate this reliability are not true in all situations.) Using all course grades that made up one semester's GPA, Ramist, Lewis, and McCamley, in the paper "Implications of Using Freshman GPA as the Criterion for the Predictive Validity of the SAT," estimated the reliability of one course grade as .35. In the example that follows, .35 is used as an upper bound for the reliability of a course grade.

If this reliability is used as an upper bound for a course grade and .90 is the maximum reliability for most CPTs, the maximum correlation possible between a CPT score and course grade is calculated by multiplying the square roots of these two reliabilities together. Here, the square root of .35 is .59 and the square root of .90 is .95. Multiplying .95 times .59 gives .56. The maximum correlation, then, between a course grade with a reliability of .35 and a test with a reliability of .90 is .56.

Using the estimate provided by Ramist, Lewis, and McCamley as an upper bound, one can easily see how instead of the expected range of 0 to 1 for correlations, the actual range would be 0 to .56. In this context, correlations of .40 and .50 are then relatively large correlations. In fact, the median correlations between the CPTs and end-of-term grades cited in the 1988 *Computerized Placement Tests Coordinator's Notebook* are from .29 to .53.

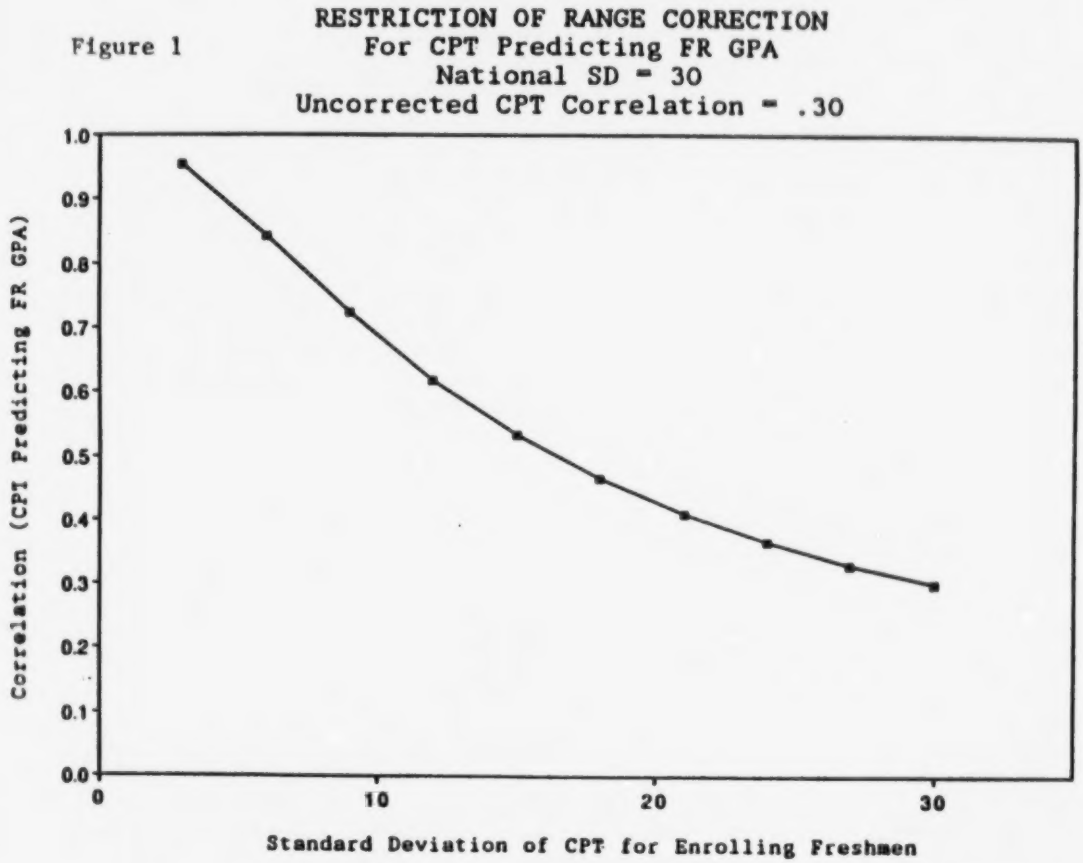
Another factor that depresses correlations between course grades and the CPTs is the restricted range of ability present in the group of students taking a course. Correlations are directly influenced by the variability in the correlated variables. When correlations are based only on the students in the course, the range of scores on the CPTs is seldom representative of all the students who take the CPTs. In most instances, courses are offered only to students with a certain course background. Students who lack the course prerequisites probably have test scores below those obtained by students in the course and students

who have been given credit for the course probably would score higher than students in the course.

Figure 1 shows how the restriction of range correction would affect the correlation obtained in a study of the relationship between CPTs and course grades. The horizontal axis displays the standard deviation of the group of students upon which the correlation was calculated. The vertical axis displays values of the correlation coefficient. The curved line in the graph shows how the obtained correlation changes when the correlation is corrected for the variability of the group of students used to calculate the correlation. The obtained uncorrected correlation was .30 for the example provided in this attachment, and the graph illustrates how this correlation would increase, when corrected, as a function of the standard deviation of scores of the enrolling freshmen. For instance, if the standard deviation of scores of the enrolling freshmen was only 12, while the national standard deviation was 30, then the obtained correlation of .30 would increase to .62 when corrected for restriction of range.

Together, all of the factors noted above – course grade multidimensionality, the reliability of the course grades, and the restricted range of abilities of students in courses – act to depress the correlations between the CPTs scores and course grades. Included is another graph that could be used to illustrate the combined effect of the last two factors. The pie chart (Figure 2) shows an example of the combined effects of reliability and range restriction in the percent of variance (squared

Figure 1



correlation) that is explainable by the predictor – here a CPT score or scores.

The pie chart divides the variance of freshmen GPA into predicted, additional predicted if no restriction of range, impossible to predict, and potential improvement in prediction categories. In this example, the reliability of freshmen GPA is .75, and as a result, 25 percent (1 to .75) of the variance of freshmen grade point average is impossible to predict due to the unreliability of course grades. So, the actual range for the possible explained variance in this example is 0 to .75. The correlation of the CPT score with freshmen GPA is .563. By squaring the correlation of .563, we obtain the percent of variance in freshmen GPA explained, 32 percent, by the CPT score. However, when this correlation, .563, is corrected for range restriction, the correlation increases to .684. As a result, another 15 percent of the variance, for a total of 47 percent, is predicted when there is no restriction of range.

In practice, it is usually not possible to estimate the reliability of a course grade (only a lower limit can be easily estimated). The correction for range restriction is easier to estimate, but one needs to define the population used as the reference group in the correction.

Mark Pomplun is a Measurement Statistician with Educational Testing Service.

SAMPLE COLLEGE
PREDICTION PIE CHART

Figure 2

ALL STUDENTS
TOTAL GROUP
CRITERION: FR GPA
PREDICTORS: CPT

Reliability of FR GPA: .75

Impossible to predict: 1 - .75 = .25

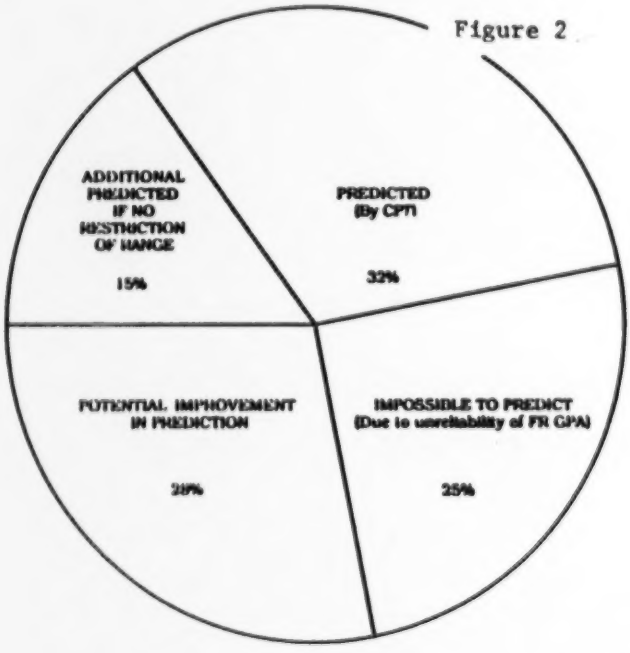
Correlation: .563

Predicted: $(.563)^2 = .32$

Correlation without restriction of range: .684

Predicted without restriction of range: $(.684)^2 = .47$

Potential improvement in prediction: .75 - .47 = .28



Computerized Placement Tests

Institutional Users as of October 1991

Aims Community College, Greeley, CO

American River, Sacramento, CA

Anson Community College, Ansonville, NC

Austin Community College, Austin, TX

Bermuda College, Devonshire, Bermuda

Boise State University, Boise, ID

Brevard Community College, Melbourne, FL

Bristol Community College, Fall River, MA

Brookdale Community College, Lincroft, NJ

Broward Community College, Coconut Creek, FL

Bunker Hill Community College, Boston, MA

California State University - Dominguez Hills, Carson, CA

Camden Community College, Blackwood, NJ

Cape Cod Community College, West Barnstable, MA

Catawba College, Salisbury, NC

Catonsville Community College, Catonsville, MD

Central Florida Community College, Ocala, FL

Central Piedmont Community College, Charlotte, NC

Central Washington University, Ellensburg, WA

Clackamas Community College, Oregon City, OR

Clayton State College, Morrow, GA

College of DuPage, Glen Ellyn, IL

College of Lake County, Grayslake, IL

College of Notre Dame, Belmont, CA

College of the Desert, Palm Desert, CA

College of the Siskiyous, Weed, CA

Coppin State College, Baltimore, MD

Crafton Hills College, Colton, CA

De Anza College, Cupertino, CA

Dixie College, St. George, UT

Dundalk Community College, Baltimore, MD

Durham College, Oshawa, Ontario Canada

Edison Community College, Fort Myers, FL

El Camino College, Torrance, CA

Fairmont State College, Fairmont, WV

Fayetteville State University, Fayetteville, NC

Feather River College, Quincy, CA

Florence-Darlington Technical College, Florence, SC

Florida Community College, Jacksonville, FL

Glendale Community College, Glendale, CA

Gogebic Community College, Ironwood, MI

Grays Harbor College, Aberdeen, WA

Gulf Coast Community College, Panama City, FL

Harding University, Searcy, AR

Harrisburg Area Community College, Harrisburg, PA

Horry Georgetown Technical College, Conway, SC

Houston Bapist University, Houston, TX

Humber College North Campus, Etobicoke Ontario, Canada

Jefferson Community College, Watertown, NY

John Tyler Community College, Chester, VA

Kings River Community College, Reedley, CA

Laboure College, Dorchester, MA

Lake City Community College, Lake City, FL

Lake-Sumter Community College, Leesburg, FL

Lenoir Community College, Kinston, NC

Lethbridge Community College, Alberta, Canada

Lewis & Clark Community College, Godfrey, IL

Lincoln University, Jefferson City, MO

Linn-Benton Community College, Albany, OR

Long Island University, Southhampton, NY

Manchester Community College, Manchester, CT

Massachusetts Maritime Academy, Buzzards Bay, MA

McCormick Junior College, Chicago, IL

Mendocino College, Ukiah, CA

Merced College, Merced, CA

Mercer County Community College, Trenton, NJ

Miami-Dade Community College, Miami, FL

Mid-Michigan Community College, Harrison, MI

Mount Royal College, Calgary, Alberta, Canada

Mt. Wachussetts Community College, Gardner, MA

New Mexico State University - Grants, Grants, NM

North Greenville College, Tigerville, SC

Northeastern State University, Tahlequah, OK

Ohio University, Ironton, OH

Oklahoma City Community College, Oklahoma City, OK

Oklahoma State University, Okmulgee, OK

Palm Beach Community College, Lake Worth, FL

Palomar College, San Marcos, CA

Pasadena City College, Pasadena, CA

Peru State College, Peru, NE

Philadelphia College of the Bible, Langhorne, PA

Raritan Valley Community College, North Branch, NJ

Richmond Community College, Hamlet, NC

Salt Lake Community College, Salt Lake City, UT

San Juan College, Farmington, NM

Santa Fe Community College, Gainesville, FL

Santa Rosa Junior College, Santa Rosa, CA

Scottsdale Community College, Scottsdale, AZ

Sheridan College, Sheridan, WY

Shorter College, Rome, GA

Solano Community College, Suisun, CA

Springfield Technical College, Springfield, MA

St. Petersburg Community College, St. Petersburg, FL

Suffolk County Community College, Selden, NY

Suffolk County Community College, Brentwood, NY

Suffolk County Community College, Riverhead, NY

Sullivan County Community College, Loch Sheldrake, NY

Sumter Area Tech College, Sumter, SC

Surry Community College, Dobson, NC

Tallahassee Community College, Tallahassee, FL

Texas State Technical Institute, Waco, TX

Tri-County Technical College, Pendleton, SC

Trident Technical College, N. Charlestown, SC

Truckee Meadows, Reno, NV

Tulsa Junior College, Tulsa, OK

Tunxis Community College, Farmington, CT

U.S. Merchant Marine Academy, Kings Point, NY

Ulster County Community College, Stoneridge, NY

University of Hawaii - West Oahu, Pearl City, HI

University of Kentucky - Owensboro, Owensboro KY

University of Maine - Orno, Orno, ME

University of Maine - Presque Isle, Presque Isle, ME

University of Minnesota - Crookston, Crookston, MN

University of New Mexico, Los Alamos, NM
University of Texas - El Paso, El Paso, TX
Victor Valley College, Victorville, CA
Vincennes University, Vincennes, IN
Wayne State University, Wayne, NE
Western Iowa Technical College, Sioux City, IA
Westfield State College, Westfield, MA
Wright State University, Celina, OH
Yuba College, Marysville, CA

NOTES

Send for your examination sets. Fill out and mail the reply card today!

Use this innovative system to stay out in front of the total placement process.

Computerized Placement Tests

ACCUPLACER™

Yes. I want to know more about Accuplacer, the Computerized Placement Tests, and the Computerized Placement Management Software. I understand that I will be billed for the examination set(s) at the price(s) listed below and that the fees will be credited toward my first order. I agree to complete and return the licensing agreement that will accompany the examination set(s), and also to return the examination set(s) at the end of the review period.

A free copy of the **ACCUPLACER** Demo Disk.
☐ 3½" disk 214648 ☐ 5¼" disk 214649 **No charge**

Computerized Placement Test Exam Set.
☐ 3½" disk 210481 ☐ 5¼" disk 210479 **$25.00 each**

Computerized Placement Management Software.
☐ 3½" disk 210483 ☐ 5¼" disk 210482 **$10.00 each**

Name _____

Title _____

Department _____

College/University _____

Address _____

City _____ State _____ Zip _____

Telephone (___) _____

FAX number (___) _____

NO POSTAGE NECESSARY IF MAILED IN THE UNITED STATES

BUSINESS REPLY MAIL
FIRST CLASS MAIL PERMIT NO 80 PRINCETON, NJ

POSTAGE WILL BE PAID BY ADDRESSEE

The College Board
Computerized Placement Tests (CPTs)
P.O. Box 6800
Princeton, NJ 08543-6901